UNDERSTANDING THE EFFECTS OF SLEEP DEPRIVATION ON EXECUTIVE FUNCTION, COMPLEX TASK PERFORMANCE AND SITUATION AWARENESS

By

Nancy Lynn Grugle

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

In

Industrial and Systems Engineering

Dr. Brian Kleiner, Chairman

Dr. Kari Babski-Reeves

Dr. Thomas Balkin

Dr. Tonya Smith-Jackson

Dr. Nancy Wesensten

March 14, 2005

Blacksburg, Virginia

Keywords: Sleep Deprivation, Executive Function, Situation Awareness

UNDERSTANDING THE EFFECTS OF SLEEP DEPRIVATION ON EXECUTIVE FUNCTION, COMPLEX TASK PERFORMANCE AND SITUATION AWARENESS

Nancy L. Grugle

ABSTRACT

Both sleep deprivation and loss of situation awareness (SA) have been cited as primary causal factors contributing to the accident and injury rate in the military and civilian sector (e.g., transportation). Despite the numerous references to both factors as causal in nature, much of the literature on the effects of sleep deprivation on executive function is anecdotal. Research has produced mixed results regarding the nature and extent of performance degradation on a variety of lower-level and executive function tasks. Similarly, although SA has been cited as a significant contributor to operational performance, there is still considerable debate over the definition and construct validity of SA. Thus, a 29-hour hour sleep deprivation study was conducted to analyze the effects of sleep deprivation on both lower-order cognitive tasks (e.g., attention and working memory) and executive function tasks (e.g., reasoning, planning, decision making, and SA). In conjunction with the sleep deprivation analysis, the relationships among lower level cognition, executive function, and situation awareness were analyzed to form hypotheses about the SA construct and its relationship to complex task performance.

Forty-eight participants were administered a series of cognitive tasks during baseline and sleep deprived testing sessions. Paired t-tests and additional post hoc analyses were conducted to determine the effects of sleep deprivation on cognition. Regression and factor analysis were used analyze the relationship among lower-order cognition, executive function, situation awareness, and complex task performance. Paired t-test results showed degraded vigilance in response to sleep deprivation, but did not indicate degraded executive function. Results of additional post-hoc analyses on executive function data indicated a trend toward degraded decision making and a trend toward increased planning errors in response to sleep deprivation. The results of the regression and factor analyses provided initial support for a dynamic, process definition of SA and illustrated the importance of considering SA as part of information processing as a whole in order to improve performance prediction. Based on the results of this

dissertation, engineering recommendations were made for developing an “ideal” SA measurement technique and improving existing SA measurement techniques. Additionally, future sleep deprivation and situation awareness research directions were suggested.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

On May 19, 2003, US Secretary of Defense Donald Rumsfeld issued a challenge to the Department of Defense (DoD) to reduce the accident and injury rate of DoD personnel by 50% within the next two years. This challenge was issued for two reasons. First, a report by the Congressional Research Service concluded that the DoD had recently achieved little progress in their attempts to reduce aviation accidents. Second, there had been a recent surge in non-aviation accidents both at home and abroad (Selinger, 2003). Both sleep deprivation and loss of situation awareness (SA) have been cited as primary causal factors contributing to the accident and injury rate in the military (Jones and Endsley, 1996; Knapp and Johnson, 1996; Taylor, 1990). However, much of the literature on the effects of sleep deprivation on executive function is anecdotal and the causal relationship between sleep deprivation, executive function, and accidents is not fully understood. Furthermore, while impaired SA has been cited as a factor in many accidents, there is still considerable debate over the definition of SA and whether SA is a distinct phenomenon. Thus, the purpose of this dissertation was twofold: 1) to determine the effects of sleep deprivation on executive functioning and complex task performance, and 2) to gain a better understanding of the interrelationship between the SA construct, executive function, and complex task performance.

As a first step toward meeting the Secretary's challenge, it is important to understand the nature of the modern military operational environment, and to clarify those aspects of the environment that are likely to contribute to errors and accidents. The military operational environment has always been psychologically and emotionally stressful, physically challenging, and cognitively demanding. However, largely as a function of recent technological advances, military operations over the past few decades have also increasingly become continuous, 24-hour-per-day endeavors. In fact, such sustained operations characterized by prolonged work, sleep deprivation, and underfeeding (Lieberman, Tharion, Shukitt-Hale, Speckman, and Tulley, 2002; Nindl et al., 2002) are currently becoming the norm rather than the exception as part of a military-wide reorganization of combat and institutional organizations in response to the current and projected security environment post September 11, 2001 (*The Way Ahead*, 2003).

Within the modern operational environment, the safety and effectiveness of all warfighters (e.g., pilots, ground soldiers, support units) depends on sustainment of operationally-relevant mental abilities (i.e., those abilities that facilitate recognition of dangers as well as targets of opportunity, that enable timely and correct decision-making, and that result in swift and coordinated action on the battlefield). However, the sleep deprivation that invariably results from participation in continuous operations may be impinging on those mental abilities most critical to maintaining safety and effectiveness in the operational environment.

The possible role of sleep deprivation in military accidents has come to national attention since military actions began in Afghanistan and Iraq. Significant news coverage publicized the issue of sleep deprivation in combat when two sleep-deprived Air Force pilots, after taking d-amphetamine, inadvertently bombed four Canadian soldiers in Afghanistan (*Need for speed: Did amphetamines play a role in Afghanistan friendly fire incident?*, 2002). More recently, the Los Angeles Times (*The development of situation awareness measures in ATM systems*, 2003) reported that a convoy of armored vehicles traveling from Kuwait to Iraq were caught in a traffic jam in the middle of the desert because large numbers of drivers fell asleep at the wheel. Furthermore, journalists embedded with that convoy reported exhaustion among the soldiers as well as military vehicles drifting off route and vanishing in the sandstorms as the driver slept (Jay, 2003).

The issue of sleep deprivation and its role in accidents is not new. A lack of sleep in military operations has been a known factor in catastrophic failures, accidents, and friendly-fire incidents (Belenky et al., 1994). Fatigue and sleep deprivation are a problem for not only the front-line troops, but also the support personnel in logistics, maintenance, air traffic control, and radar. During a deployment that requires 24-hour activity, human error as a result of sleep deprivation can cause friendly-fire incidents as well as aviation and driving accidents (Moore-Ede, 2003). During peacetime operations, fatigue contributes to the abandonment of safety practices which causes training injuries, maintenance mistakes, and aviation and motor vehicle accidents. The results of these accidents are reduced unit readiness, degraded effectiveness, and disrupted operations (*Need for speed: Did amphetamines play a role in Afghanistan friendly fire incident?*, 2002).

The military has examined the contribution of sleep deprivation to accident rates in the past. The U.S. Air Force estimated that fatigue costs over $54 million in total Class A accidents each year. Between 1977 and 1990, the Air Force attributed approximately 25% of the night tactical fighter Class A accidents to fatigue. During those same years, the Navy estimated that 12.2% of total Class A mishaps were due to aircrew fatigue. The U.S. Army Safety Center reported that 4% of the Army's total mishaps were fatigue-related between 1990 and 1999 (J. A. Caldwell and Gilreath, 2002). Approximately 9% of wheeled-vehicle crashes that resulted in injury or death during the Gulf War were due to driver drowsiness and fatigue (Peters et al., 1999). The precise contribution of sleep deprivation to friendly-fire incidents is not known. However, the critical functions necessary to prevent fratricide (e.g., awareness of and orientation to friendly and enemy troops, target designation and tracking, and fire control) deteriorate rapidly when soldiers are sleep deprived. It is possible, if not likely, that sleep deprivation contributes to friendly-fire incidents. To illustrate the potential significance of sleep deprivation in friendly-fire accident rates, a study on Operation Desert Storm indicated that 24% of soldiers killed in action and 15% of soldiers wounded in action were the result of friendly-fire. In addition, 77% of all allied combat vehicles lost in the conflict were destroyed by U.S. forces (Steinweg, 1995). Although there is no data on the role of sleep deprivation in these incidents, the potential contribution of sleep deprivation is potentially significant.

Much of the research and statistics on accidents related to sleep deprivation and fatigue involve civilian accident data. Many of the causal factors of civilian driving accidents are similar to those in the military (e.g. driver sleepiness, boredom, and monotony), with the potential exception of driving during combat. The following section specifically addresses the contribution of sleep deprivation to driving accidents and discusses the causal factors of these accidents.

In a joint study by the Federal Highway Administration's Human Factors Laboratory and the Walter Reed Army Institute of Research, the effects of increasing sleep deprivation on simulated driving performance and accident rates under controlled conditions was investigated (Barfield, Rosenberg, and Furness, 1995). Twelve participants drove a high-fidelity simulator for 40 minutes per day for four days under increasingly sleep-deprived conditions. Table 1 shows the level of sleep deprivation on each day of testing.

Table 1. Sleep Deprivation Conditions on Consecutive Test Days (adapted from Peters et al., 1999)

Test Day	Hours of Continuous Wakefulness	Level of Deprivation
1	9	None
2	12	4 hours of sleep
3	36	No sleep for one day
4	60	No sleep for two days

Accidents increased immediately following a night restricted to only four hours of sleep and continued to increase with continuing sleep deprivation. Performance on days 3 and 4 showed large numbers of crashes and lane excursions. After 60 hours of sleep deprivation, participants reported difficulty concentrating as well as visual misperceptions and distracting thoughts.

A study by De Valck and Cluydts (2001) showed similar results. They found that drivers who obtained only 4.5 hours of sleep the night before testing showed increased lane drifting and increased speed deviations compared to a control group that had obtained 7.5 hours of sleep. There was no significant difference in the number of accidents between the two groups. However, the study only tested the effects of one night of sleep restriction on performance.

Connor et al. (2002) conducted a population-based control study to determine the contribution of driver sleepiness to car crash injuries. A total of 615 crashes that resulted in injuries or fatalities were studied by interviewing drivers either treated in the emergency room or admitted to the hospital. In the case of fatalities, the coroner was consulted. Strong correlations were found between driver sleepiness as measured by the Stanford Sleepiness Scale (score of 4 or higher) and risk of injury resulting from a crash. Drivers with less than five hours of sleep in the 24 hours prior to the crash had a significantly higher risk of a crash resulting in injury than drivers who had more than five hours of sleep.

In addition to driver sleepiness, other factors that contribute to fatigue-induced driving accidents include the cognitive decrements associated with sleep deprivation, the monotony of driving, the time of day, and age. Sleep deprivation results in degraded perception and vigilance, as well as increased distractibility and confusion. Each of these factors affects a person's ability to safely operate a vehicle and may play a central role in driving accidents (Barfield et al., 1995).

Harrison and Horne (1999) reported that both time of day and age play a role in accidents attributed to driver fatigue. Sleepiness-related driving accidents peak between 0200-0600 and 1400-1600. For example, drivers are 20 times more likely to fall asleep while driving at 0600 than at 1000. The authors also reported that drivers under 30 years of age were more likely to have sleepiness-related driving accidents than older drivers. This trend could be attributed to increased susceptibility to sleep loss or to young drivers' tendency to ignore sleepiness and continue driving.

In addition to contributing to driving accidents, sleep deprivation has been a causal factor in many aviation accidents. Well-known aviation accidents caused by sleep deprivation and fatigue include the crash of Korean Air flight 801 in August 1997, the DC-8 Accident at Guantanamo Cuba Naval Base, and the near-crash of China Airlines Flight 006 in 1985 (Caldwell and Gilreath, 2002). Caldwell (1997) indicated that while many of the human errors that account for over half of all aviation accidents are likely the result of fatigue in some way (e.g., pilot inattentiveness and failure to respond to critical information), fatigue is not usually cited as the primary causal factor. Importantly, frequently cited causal factors such as complacency, distraction, sensory illusions, and inadequate resource utilization can all result from aircrew sleepiness and fatigue. In an investigation into F-16 mishaps in the U.S. Air Force from 1975-1993, researchers found that fatigue was either definitely, probably, or possibly a cause in approximately 3.8% of all mishaps, while channelized attention, loss of situation awareness, distraction, visual illusion or reduced cues, and complacency accounted for a combined 35.8% of all mishaps (Knapp and Johnson, 1996). All of these factors can be caused by or exacerbated by sleep deprivation.

Caldwell and Gilreath (2002) conducted a survey of 401 Army aviators and aircrew members to determine whether fatigue was an issue in Army Aviation. The most noteworthy responses were in regard to the following questions: (1) Have you ever had to fly…when you were so drowsy you felt you could easily fall asleep? (2) Have you ever dozed off while flying/in the cockpit (even just a momentary, nonthreatening nodding off)? (3) Is there a widespread problem in the military aviation community with flying, or performing other critical aviation duties, while too tired? Seventy-two percent of pilots and 85% of nonpilots indicated that they had flown when they were so drowsy they felt they could fall asleep, and even more importantly, 45% of pilots and 46% of nonpilots admitted they had actually fallen asleep while

flying or in the cockpit. Furthermore, 73.4% of pilots and 86.7% of nonpilots felt there was a widespread problem in the military aviation community of performing duties while too tired. Considering the number of responses that indicated sleep deprivation and fatigue are major issues in the U.S. Army Aviation community, it is not surprising that sleep deprivation is cited as a major contributing factor to aviation accidents in the military.

Sleep deprivation potentially contributes to aviation accidents in several ways. Proficient performance on the modern battlefield involves vigilance, reaction time, situation awareness, memory, and decision making (Lieberman et al., 2002). In general, sleep loss results in lapsing, cognitive slowing, memory impairment, and vigilance decrease (Himashree, Banerjee, and Selvamurthy, 2002). Impaired working memory, verbal fluency, logical reasoning, decision making, and judgment may also result from sleep deprivation (Balkin et al., 2000). With respect to mood, sleep-deprived people may experience irritability, forgetfulness, and aversion to effort (Wiegmann, Stanny, McKay, Neri, and McCardie, 1996). Thus, sleep deprivation degrades critical cognitive processes required for performance in the operational environment. In addition to the general degradation of cognitive performance and mood that occurs with prolonged sleep deprivation, sleep loss in the aviation environment presents specific problems. Pilots may lose the ability to perceive and integrate important information due to attentional narrowing. Performance may become less consistent and the ability to follow procedures may be degraded. In addition, involuntary lapses into may sleep occur (Caldwell, 1997). Despite a general consensus on the likely cognitive effects of sleep deprivation, many of the studies mentioned above are making hypotheses about the link between sleep deprivation and executive function. In fact, there have been few scientific investigations into the effects of sleep deprivation on executive function. Therefore, a critical fist step in reducing the accident and injury rate is to investigate the effects of sleep deprivation on higher order cognitive function.

Moreover, accidents caused by sleep deprivation are likely not the result of decreased alertness alone. The interaction of sleep deprivation with environmental conditions and complex task requirements may produce a combined effect that contributes significantly to the accident rate in the military. For example, degraded situation awareness (SA) has been cited as causal factor in numerous aviation mishaps. Impaired SA was cited as the primary causal factor in 20 out of 41 U.S. Air Force operator error accidents in 1984 and a 1991 review of 175 military aviation mishaps (Jones and Endsley, 1996; Taylor, 1990). SA deficits were cited as the second

leading cause of F-16 Class A mishaps in the U.S. Air Force between 1975 and 1993 (Knapp and Johnson, 1996). Moreover, a Department of the Army report on fratricide stated that degraded SA was one of the two main causes of fratricide in the Gulf War (Steinweg, 1995). An investigation into aircraft accidents for major civilian airlines between 1989 and 1992 indicated that human error was the most prevalent factor in 71% of the accidents. Of those attributed to human error, the primary factor was degraded SA. Moreover, almost 1/3 of the accidents attributed to SA had a second contributing factor--fatigue (Jones and Endsley, 1996). A study on helicopter rotor blade injuries emphasized that safety training is not sufficient to prevent accidents, maintaining SA around the aircraft was the critical element to preventing injury (Crowley and Geyer, 1993). It is hypothesized that sleep deprivation degrades the higher, more complex cognitive processing ability required for SA. For example, sleep-deprived individuals may not have the ability to integrate multiple sources of information into a clear and accurate representation of the tactical situation (Jay, 2003). Therefore, the very cognitive processes that comprise SA and allow pilots to function effectively on the battlefield may be the ones that are degraded by sleep deprivation. Although sleep deprivation and loss of SA have been cited as causal factors in military accidents and injuries, few scientific studies have been conducted to determine the effects of sleep deprivation on executive function in a dynamic, complex task environment. Thus, the underlying causes of accidents attributed to sleep deprivation are not fully understood. Moreover, SA has yet to be operationally defined or justified as a construct separate from higher order cognition and worthy of investigation as a distinct phenomenon. Therefore, in order to design successful engineering interventions for personnel, technology, and the military organization, more research is needed to understand the effects of sleep deprivation on lengthy, complex task performance and to simulate an operational setting such that results can be generalized to military operations in the field. Additionally, before the effects of sleep deprivation on SA can be determined, the construct validity of SA must be investigated to better understand the relationship between executive function and SA.

An example of such an engineering intervention based on sleep deprivation research is the Walter Reed Army Institute of Research (WRAIR) sleep management system (SMS). The SMS consist of six components (Wesensten, 2003):

1. A wrist-worn activity monitor (actigraph) to continuously record wrist movements

2. An algorithm to score sleep vs. wake time from recorded activity
3. A mathematical sleep/performance model to predict performance based on scored sleep
4. Recommendations for stimulant usage to sustain cognitive performance when sleep is not possible
5. Recommendations for the use of sleep-inducing agents to induce recuperative sleep and, when necessary, the use of rapid reawakening agents to restore alertness and performance
6. Guidelines and doctrine for the SMS to manage sleep, alertness, and performance

The actigraph is an objective and unobtrusive measure of a soldier's sleep duration, continuity and timing during sustained operations. The sleep/performance model is based on laboratory studies that investigated the effects of sleep deprivation on cognition (e.g., serial add/subtract). The model predicts current and future performance based on the prediction equation and sleep recorded by the actigraph. When no sleep is possible, stimulants may be used to counteract the performance-degrading effects of sleep deprivation when no sleep is possible. Sleep-inducing agents may be used to help the soldier fall asleep quickly when necessary and rapid reawakening agents have the potential to restore alertness quickly when the soldier must be awakened. In addition, the SMS provides guidelines for sleep management so the system can be effectively implemented in the field. The six SMS elements work together to manage, monitor and predict a soldier's sleep such that safety and performance are maximized on the battlefield (Belenky, 1997).

In order to successfully reduce the number of accidents and injuries, it is critical to understand that causal factors such as sleep deprivation and SA are interrelated and cannot be studied in isolation from personnel, technology, or the battlefield environment. It is the understanding of this interaction and the prescription of appropriate interventions based on that understanding that will ultimately lead to a safer and more efficient military. Thus, both research and engineering solutions should take a macroergonomic approach. Research should be conducted to better understand the effects of sleep deprivation on executive function, including the relationship between executive function, complex task performance, and the construct of SA. Results of the research should then be considered within the context of the entire military system to include personnel, technology, and the operational environment using a macroergonomic approach and data incorporated appropriately into the U.S. Army's Human Systems Integration program.

The U.S. Army developed a program of Manpower and Personnel Integration (MANPRINT) in 1986 in response to increasing frequency and severity of accidents involving human error. Booher (2003) describes the HSI program in detail. The program was adopted by DoD and is currently referred to as Human Systems Integration (HSI) in that organization. The goal of HSI is to effectively integrate human factors into all aspects of military system design so that soldier performance and productivity is improved while accidents and injuries are reduced. To ensure a systems approach, the program encompasses many areas of the Army including Army Material Command, Training and Doctrine Command, Office of the Surgeon General, Army Safety Center, Army Research Institute, and Human Engineering Laboratory. The human-related technologies and disciplines include the following seven domains:

- Manpower
- Personnel
- Training
- Human Factors Engineering
- System Safety
- Health Hazards
- Soldier Survivability

Manpower refers to the number of men and women required and available to operate and maintain military systems. Personnel is the physical and cognitive human capabilities and characteristics required for training, operation, maintenance, and sustainment of materiel and information systems. Training refers to the education required to provide personnel with adequate skills, knowledge, and attitudes for job performance (Kleiner and Booher, 2003). Human factors engineering is the integration of human capabilities and limitations into system definition, design, development, and evaluation. Systems safety includes design and operating characteristics for minimizing potential human and system errors that may cause injury or accidents. Health hazards involve the system design and operating features that create risk of bodily injury or death including biological hazards, chemical exposure, etc. Soldier survivability defines the system attributes that can reduce fratricide and detectability, and minimize system damage, soldier injury, and fatigue. When taken together, integration of each of these domains into a single HSI process ensures the successful integration of people with technology and organizations across all systems (Belenky, 1997).

The scope of this dissertation falls within the soldier survivability domain of the HSI program. Booher (2003) identified the quantification of human parameters as one of the 10 principles critical to effective HSI. These parameters include physical and physiological characteristics, cognitive ability, and social skills as well as the levels of stressors that may affect soldier safety and health. This dissertation contributed to soldier survivability data by quantifying the effects of sleep deprivation on executive function and complex task performance. More specifically, the data can potentially be applied to the Walter Reed Army Institute of Research (WRAIR) Sleep Management System (SMS) as an extension to existing modules to better predict soldier performance under sleep-deprived conditions. This dissertation provided additional data to facilitate the evolution of the SMS with respect to predicting and facilitating higher order and complex task performance under sleep-deprived conditions. Furthermore, it aids in the development of effective countermeasures for operational performance on the battlefield. Countermeasures might include performance prediction indicators or sleep/wake schedules, for example.

The investigation of SA contributed to the soldier survivability domain as an initial investigation into the relationship among SA, executive function, and sleep deprivation. Although SA has been cited as a causal factor in many military accidents, the SA construct has not been clearly defined, validated, or sufficiently measured. Measuring the effects of sleep deprivation on an ill-defined and insufficiently validated construct may provide insight into overall performance deficits on several outcome metrics, but it will provide little or no diagnostic capability for isolating and understanding the underlying causes of any degradation. A better approach is to first understand the relationship between the higher order cognitive functions required for SA and the effects of sleep deprivation on those functions. Therefore, examining SA with respect to executive function and complex task performance will provide data to either support or refute the SA construct as a phenomenon separate from lower order and higher order cognitive function. In addition, data will quantify which aspects of SA are more degraded by sleep deprivation than others. Data can then be used to design engineering solutions and interventions that can be applied to technology, personnel, and the environment.

Although the scope of the dissertation was applied narrowly in the soldier survivability domain, when considered within the entire HSI framework, the dissertation facilitated the integration of knowledge of human capabilities and limitations into the military organization and

feed into each HSI domain. Interventions to minimize the effects of sleep deprivation on performance can be applied to technology in the form of a sleep management system, personnel in the form of stimulants, and the organization in the form of sleep management policies.

1.1 Research Questions

This dissertation addressed the following research questions related to the effects of sleep deprivation on executive functioning and complex task performance:

- Which executive functions were degraded by sleep deprivation?
- How was complex task performance differentially affected by sleep deprivation as compared to performance on individual tests of lower and higher order cognition?
- How did sleep-deprived performance on a lengthy complex task change over time? For example, was complex task performance be maintained for 60 minutes under sleep-deprived conditions?
- Was performance on any individual test of higher order cognition a predictor of complex task performance?
- What was the relationship between perceived workload and complex task performance?

The following research questions address the investigation of the SA construct:

- How and to what extent did executive function relate to performance on a complex task hypothesized to require SA for good performance?
- Did the relationship between higher order cognitive functions hypothesized to comprise SA change under sleep-deprived conditions? In other words, were some aspects of cognition more important to performance under sleep-deprived conditions?

1.2 Research Model

To investigate each of the research questions above, a 29-hour laboratory sleep deprivation study was conducted during which participants were tested on several individual executive tasks as well as a complex, interdependent task hypothesized to require SA for good performance. The single study was methodologically broken into two segments.

One segment of the study analyzed the effects of sleep deprivation on executive function by comparing baseline and sleep-deprived performance on several cognitive tests. This segment used a hypothesis-testing approach to determine whether executive functions were significantly

degraded by sleep deprivation. The second segment of the study involved collecting non-sleep deprived performance measurements on all executive tasks and a 60-minute dynamic, complex task hypothesized to require SA for good performance. Data from this segment was used to analyze the relationship between executive function and SA. This analysis took an exploratory approach to better understand the relationship between executive function and the SA construct and to develop hypotheses about the nature of the SA construct. Figure 1 shows the inputs and outputs for each segment of the proposed study.

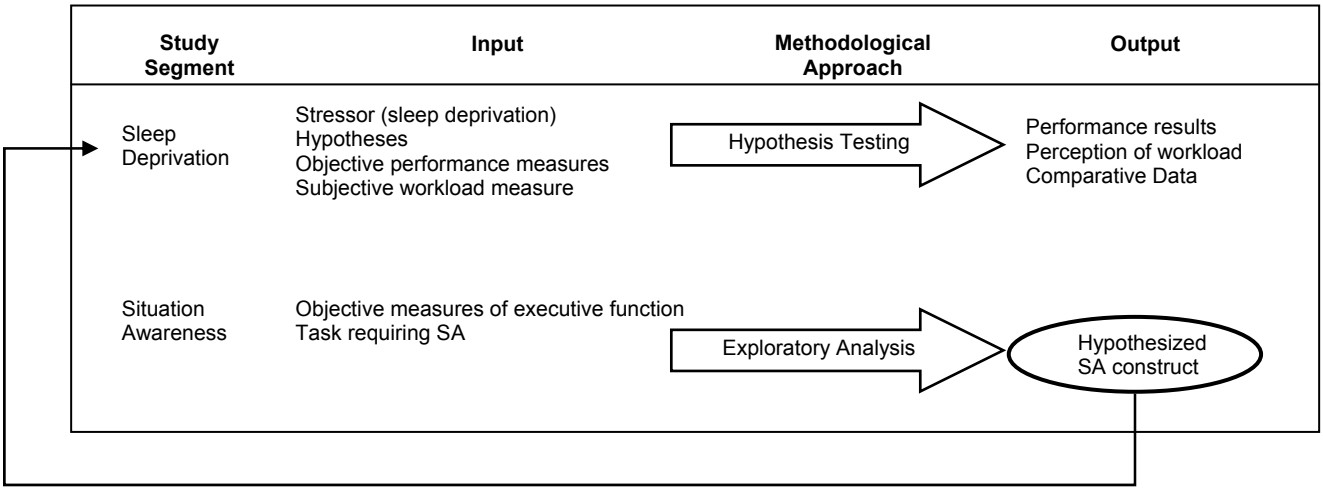


Figure 1. Input/Output Research Model

While there were similar inputs to the two study segments, the distinct methodological approaches result in different outputs. The sleep deprivation research resulted in both subjective and objective performance data. The SA research resulted in a proposed definition of SA as well as hypotheses about the nature of the SA construct. Figure 2 shows how both of these research segments fit within the macroergonomic framework.

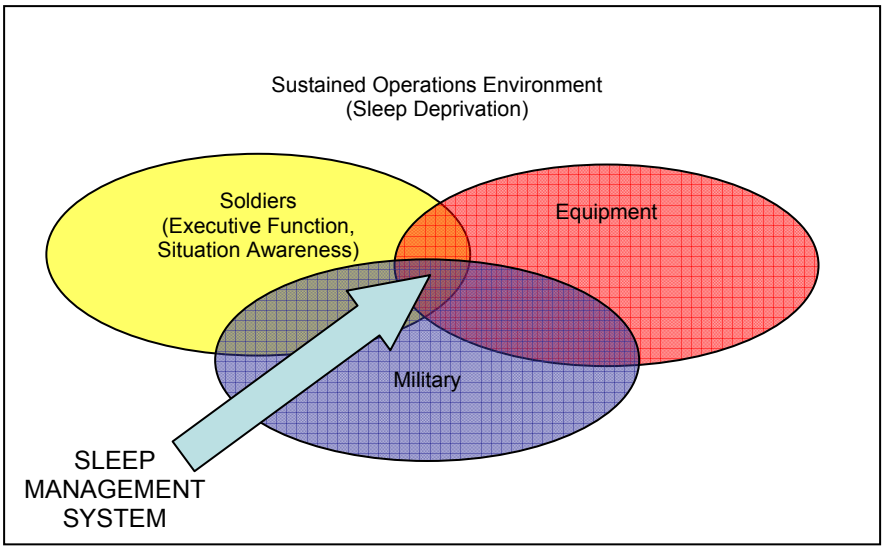


Figure 2. Macroergonomic framework for research

Both executive function and SA are personnel (soldier) factors that are characterized by individual differences in cognitive ability. The sustained operations environment includes stressors such as sleep deprivation that form the operational demands in which personnel, technology, and the organization function. The sleep management system is centered at the intersection of technology, soldiers, the military, and the sustained operations environment. Although the research variables of interest (i.e., executive function, situation awareness, and sleep deprivation) fall within only two of the four components, the results of this research will feed into all areas of the sociotechnical system to develop technological interventions, personnel performance monitoring, and sleep schedules, for example.

1.3 Hypotheses

Based on previous literature on the effects of prefrontal cortex (PFC) damage and sleep deprivation on cognitive function, this dissertation made several hypotheses about the effects of sleep deprivation on executive functions. In a sleep-deprived environment, results of studies have shown that procedural actions are typically not degraded, but that need-driven and opportunity-driven actions may be affected by inflexible thinking and delayed decision making as a result of attentional narrowing (Herscovitch, Stuss, and Broughton, 1980). Furthermore, results have indicated that the perception and integration of new, but routine or expected information was not affected by sleep loss, but that flexible responses to that information were

affected (Harrison and Horne, 1999). Therefore, this dissertation hypothesized that performance on tasks requiring routine or rule-based decision making (e.g., deductive reasoning task) would not be degraded by sleep deprivation. Conversely, performance on tasks requiring flexible thinking (e.g., inductive reasoning) and divergent decision making (e.g., Iowa gambling task), performance would be degraded by sleep deprivation.

PFC-damaged patients have demonstrated degraded ability to plan and an inability to appreciate the future consequences of action (Fuster, 1989). Thus, this dissertation hypothesized that sleep-deprived participants would show degraded performance on planning tasks (e.g., Maze Tracing Test).

Moreover, this dissertation hypothesized that as sleep loss increased, the contribution of executive functions to maintaining SA (as measured by performance on a complex task requiring SA) would change. In other words, the participant would be able to maintain some executive functions better than others while sleep-deprived and would use those functions to compensate for any degradation in the others.

It was also hypothesized that no single measure of lower or higher order cognition (e.g., vigilance) would significantly predict complex task performance. Rather, degraded higher order cognition would play a larger role in understanding degraded complex task performance. And finally, this dissertation hypothesized that sleep-deprived performance on a complex task would not be maintained for 60 minutes, regardless of the level of stimulation or participant motivation.

This dissertation further hypothesized that many of the executive functions being tested would load highly onto the SA construct. Therefore, this dissertation predicted that results would support the proposition that SA is not a construct separate from executive function, decision, and action. Rather, SA is just another term for executive functioning under complex task conditions. And finally, this dissertation hypothesized that there would be a significant relationship between workload ratings and complex task performance although it will not necessarily be causal or predictive in nature.

1.4 Study Limitations

Because the proposed study was a laboratory experiment, there were some limitations with respect to generalizability to a sustained operations environment. These are discussed from a sociotechnical systems perspective. The operational environment has multiple potential

stressors that can contribute to reduced performance. As reported by Lieberman et al. (2002), few studies have looked at the effects of using caffeine or other stimulants in environments where multiple stressors such as extreme temperatures, intense physical activity, and psychological stress are simultaneously present with sleep loss. Because a laboratory environment cannot adequately simulate all of the stressors that a soldier in a sustained operations environment experiences, it is critical to consider the possibility of an additive effect on performance.

The impact of technology on performance has been studied in an operational setting during an experiment at Fort Benning in October-December 1996. Subjective results of an after action review indicated that the Land Warrior system enhanced situational awareness, reduced soldier workload, and improved communications (Gilmore, 1997). The interaction of complex technology and sleep deprivation was not systematically studied, however, Lieberman, et al. (2002) hypothesized that as the cognitive complexity required to operate effectively on the battlefield increases with technological advances, the cognitive performance decrements due to battlefield fatigue and other stressors will be magnified. Because this study used pen and pencil cognitive tests as well as simple computerized tests, the impact of using complex technology under sleep-deprived conditions was not be addressed.

With respect to a sleep management system, relevant psychosocial factors affecting the personnel subsystem include intrinsic factors such as cognitive ability, individual differences in susceptibility to performance effects of sleep loss and stimulants, response to stressors, and variability in mood. Although the demographic factors of military personnel cannot be manipulated within a sustained operations environment, these factors may affect the management of sleep and stimulants. For example, age and gender may influence individual sleep need and should be considered when using chemical intervention as part of the sleep management system. The psychosocial factors within the personnel subsystem may have the largest impact on the development of an effective sleep management system. Therefore, while these factors can be studied in the laboratory, engineering interventions must consider the real-world implications of personnel factors on design.

Chapter 2 provides the physiological and cognitive foundation for linking brain function, sleep, executive function, and situation awareness. First, the physiological basis for executive function is described. The recuperative effects of sleep and the degrading effects of sleep

deprivation on executive function are then discussed with a focus on the frontal lobes. Next, the current literature on SA is discussed within a human information processing framework. And finally, the methodological approach to studying the effects of sleep deprivation on executive function and complex task performance is discussed as well as the approach to evaluating the situation awareness construct.

Chapter 2. Literature Review

2.1 The Human Brain

The human brain has two cerebral hemispheres covered with cerebral cortex, a layer of gray matter composed of neurons and their synaptic connection (Kolb and Whishaw, 1996). The cortex consists of layers of nerve tissue varying in cell size, type, and density. With the exception of the pineal gland, the two hemispheres contain identical functional modules. Figure 3 shows the main areas of the brain including the frontal lobe, corpus callosum, parietal lobe, cerebellum, occipital lobe, brain stem, temporal lobe, and limbic system (Carter, 1998).

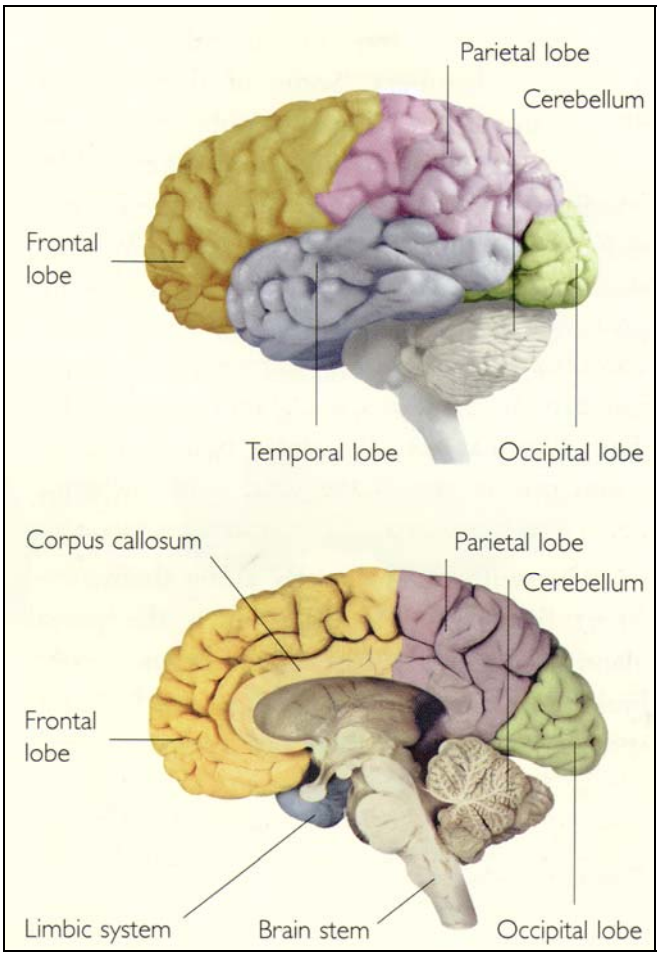


Figure 3. Functional modules of the Human Brain (Carter, 1998)

The cortex can be separated into four geographically and functionally distinct areas or lobes: the frontal, temporal, parietal, and occipital lobes. The frontal lobe is associated with

working memory, sustained attention, planning, monitoring of behavior, and temporal memory. The functions of the temporal lobe include verbal processing such as language comprehension, word recognition, and processing of sound. The parietal lobe's functions include attention, spatial perception, and working memory, skill learning, and memory retrieval. The occipital lobe is primarily responsible for visual processing including perception and manipulation of visual information as well as mental imagery (Martin, 2003). Although there is functional interaction between the brain regions, the primary area associated with executive function is the frontal lobes.

2.2 The Frontal Lobes

The frontal lobes encompass approximately one third of the cortex and contain many distinct functions (Martin, 2003). There are three main areas of the frontal lobes: the motor cortex, the premotor cortex, and the prefrontal cortex. The motor and premotor areas are part of a system that controls movements. The motor cortex executes individual movements and the premotor cortex selects the movements. The prefrontal cortex performs the cognitive processing prior to movement or action (Kolb and Whishaw, 1996).

Much of what is known about the function of the frontal lobes has been determined by analyzing cognitive deficits in patients with damage to that area of the brain. Unfortunately, in cases of lesions and brain trauma, there may be damage to other brain regions, irritation of other areas of the brain as a result of the trauma, and irregularities across lesions that make comparisons between individuals and generalizations difficult. Other more useful information has been gained from investigations of prefrontal damage as a result of frontal lobectomies, discrete trauma, and tumors. A third source of data is degenerative processes such as Alzheimer's disease and dementia which typically affects the frontal lobes (Fuster, 1989). Based on information from these sources, there is general agreement about the symptoms of prefrontal damage. These include disorders of attention and perception, motility, and temporal integration. The severity and the symptoms vary in frontal patients based on the location and extent of the damage (Fuster, 1989); however, damage to specific areas in the frontal lobes produces distinct impairments (Goldberg, 2001). Table 2 shows several frontal lobe syndromes and the corresponding lesions locations.

Table 2. Symptoms of Frontal Lobe Damage (adapted from Kolb and Whishaw, 1996)

Most Probable Symptom	**Lesion Site**
Loss of divergent thinking	
Reduced spontaneity	Orbital
Poor strategy formation	Dorsolateral
Environmental control of behavior	
Poor response inhibition	Prefrontal
Risk taking and rule breaking	Prefrontal
Impaired associative learning	Dorsolateral
Poor temporal memory	
Poor recency memory	Dorsolateral
Poor frequency estimate	Dorsolateral
Poor self-order recall	Dorsolateral
Poor delayed response	Dorsolateral

Prefrontal disorders of attention and perception include decreased general awareness of the environment, sensory neglect, increased distractibility, degraded visual searching, diminished sustained attention, and reduced concentration (Fuster, 1989). Motility disorders include altered spontaneous movement (either significantly increased or decreased) while disorders of goal-directed motor behavior are manifested in perseveration of incorrect behavior and lack of initiative (Fuster, 1989). Disorders of temporal integration are the most distinguishing feature of frontal lobe damage. These disorders are shown as an inability to initiate and carry out new goal-directed behaviors that require deliberation, choice, and the organization of a novel sequence of behaviors. Familiar and well-rehearsed behaviors are not affected regardless of temporal length (Fuster, 1989). However, the ability of these patients to develop novel strategies for solving problems is impaired despite an apparently intact understanding of the task requirements (Kolb and Whishaw, 1996).

Within the context of temporal integration, frontal patients also suffer memory, planning, and control of interference degradations as well as increased apathy (Fuster, 1989; Goldberg, 2001). Memory is affected by PFC damage in a specific way which is closely related to attention. Frontal patients do not show deficits in acquiring, retaining, or retrieving both old and new information. However, they do show degraded ability to distinguish and attend to relevant information, and hence, cannot remember particularly relevant information. Thus, the disorder is not in a general ability to remember, but rather in a specific inability to focus attention and memorize or encode essential information for achieving goals (Fuster, 1989). The most

frequently noted frontal impairment is planning. Degraded planning is most likely a result of a frontal patient's reduced foresight and his or her apathy. Planning involves forming goals and remembering those goals throughout the process of achieving that goal. As a result, an impaired ability to remember future actions translates into impaired foresight (Goldberg, 2001). Frontal patients also experience difficulty ignoring both internal and external distractions (Kolb and Whishaw, 1996). They cannot disregard irrelevant stimuli, leading to memory decrements for relevant information (Fuster, 1989). Furthermore, frontal patients rely on external cues rather than internal knowledge to guide actions. If they are unaware of subtle changes in context due to distractions, such patients will act inappropriately. Furthermore, frontal patients commonly fail to follow instructions, ignore signals that behavior is incorrect, and continue performing incorrectly (Kolb and Whishaw, 1996).

2.3 The Prefrontal Cortex

2.3.1 Anatomy and physiology of the PFC

The prefrontal cortex (PFC) is the most important area for higher order cognitive functioning. The prefrontal region can be separated by both location and function into two main areas: the dorsolateral (DLPFC) and orbital (OFC) areas (Krawczyk, 2002). The DLPFC is associated with reasoning, comparing, and evaluating. The primary function of the DLPFC is maintaining and manipulating information within working memory. Some new research has shown that it is also responsible for the categorization of novel stimuli. The OFC is associated with reward, emotion, and adapting to the environment. The OFC processes the reward value of environmental stimuli and suppresses responses to stimuli that are no longer rewarding. It is also associated with the inhibition of motor responses (Krawczyk, 2002).

Brain imaging techniques such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and cerebral blood flow (CBF) have been used to determine which areas of the brain are activated during cognitive tasks and, hence, determine which areas of the brain perform particular cognitive functions. PET injects low activity radioactive chemicals into the bloodstream which label specific compounds such as glucose or oxygen. The chemicals then decay and give off subatomic particles known as positrons which indicate where the compounds have been metabolized and distributed in the brain. Metabolism is indicative of brain activity in different areas of the brain. fMRI uses magnetic fields and radio

waves to obtain images of brain structures. fMRI uses the magnetic properties of blood to determine where blood is flowing, indicating brain activity as often as once per second (*Introduction to brain imaging techniques and other methods*, 2003; Mathias, 1996).

Decision making and related behaviors such as reward processing, planning, inductive reasoning, and manipulating information in working memory have been shown to activate the prefrontal cortex. The OFC is responsible for a central component of decision making—reward processing. The process of seeking gains while avoiding losses has been shown to activate areas of the brain including the thalamus, striatum, subgenual cingulate gyrus, and hippocampus. Interestingly, distinct areas of the prefrontal cortex are activated during wins vs. losses (Elliot et al., 2000; Zalla et al., 1999 as cited in Krawczyk, 2002).

Additional critical decision making functions such as maintaining and manipulating information in working memory, inductive reasoning, and categorizing novel stimuli are associated with the DLPFC. Maintaining and manipulating information in working memory allows for the prioritization of competing goals, evaluation of alternatives, and integration of new information into existing mental models (Krawczyk, 2002). Deciding between alternatives based on unstructured, inductive reasoning has been shown to activate the right DLPFC, ventrolateral prefrontal, and superior parietal regions while rule-based selection among alternatives activated the left DLPFC. A study by Goldberg (2001) indicated that different areas of the brain were activated during naïve, practiced, and novel tasks.

During naïve task performance, the frontal lobes are activated; however, after the task has been practiced, frontal activity decreases. When a task similar, but not identical, to the first task is introduced, an increase in frontal activity occurs. Using a PET study, Raichle et al. (1994) found the highest blood flow in the frontal lobes during initial performance of a novel task. As the task became more familiar, blood flow in the frontal lobes almost completely disappeared. As a new task similar to the first one was introduced, blood flow to the frontal lobes increased, but did not reach the same level as with the novel task (as cited in Goldberg, 2001). Frith and Dolan (1996) reported that the prefrontal areas of the brain are activated while learning a new task because it requires thought, but that once a task has become routine and thought is no longer required, the frontal areas are no more active during task performance than during rest. Interestingly, when participants were asked to think about performing a routine task, the frontal region was again activated.

(Sanfey, Rilling, Aronson, Nystrom, and Cohen, 2003) conducted an experiment to determine which parts of the brain were associated with cognitive versus emotional decision making using the Ultimatum Game. In the two-player game, a proposer offers to split a sum of money based on the percentage that the proposer and responder will each receive. The responder can either accept or reject the offer. If the responder accepts, both players receive money. If the responder rejects the offer, neither person receives any money. Offers that the responder deems unfair force the responder to choose between an emotional reject decision and a cognitive accept decision. When the responder received unfair offers, greater activation was shown in the bilateral anterior insula (typically associated with negative emotional states), the DLPFC (typically associated with goal maintenance and executive control), and the anterior cingulate cortex than when the offer was fair. Responders that rejected unfair offers showed higher insula activation than DLPFC activation and, conversely, responders that accepted unfair offers showed higher DLPFC activation than insula activation. The conclusion drawn from the experiment was that different brain regions compete to influence decision making behavior.

Evaluative judgment, the evaluation of stimuli based on an internal scale, is another critical factor in decision making. Zysset, et al. (2002) found the evaluative judgment process to activate the anterior frontomedian cortex, the inferior precuneus, and the left inferior frontal gyrus. Results of an fMRI study indicated distinct brain regions were activated during episodic memory retrieval, semantic memory retrieval, and evaluative judgment. However, there was strong interdependence between the regions for the process as a whole. The inferior frontal gyrus was activated during evaluative judgment when selecting among competing alternatives The situation model was updated by the inferior precuneus, which was activated during episodic memory retrieval.

Using the techniques described above as well as behavioral analysis, researchers have developed a theory of neuropsychological functions related to the PFC. The next section describes the functions of the PFC as well as the decrements in cognitive performance as a result of damage to the PFC.

2.3.2 Neuropsychology of the PFC

In general, the functions associated with the PFC include planning, discriminating, decision making, directing and sustaining attention while ignoring distractions, and initiating goal-directed behavior (Kolb and Wishaw, 1985). The PFC has been associated with flexible

and innovative thinking and decision making in response to novel and unexpected information and events (Fuster, 1989).

Damage to the PFC often results in degraded decision-making. A major consequence of OFC damage is decreased appreciation for the possible negative consequences of actions (Bechara, Damasio, Damasio, and Anderson, 1994). Patients with OFC damage show increased risk-taking and impulsive behaviors, regardless of the appropriateness of the action. The DLPFC has also been indicated as the location for integration of information in working memory. In a study investigating whether the integration of relational information depends upon the PFC, patients with DLPFC damage showed significant impairment on problems that required integration of two or more related pieces of information as compared to controls. Studies on damage to the PFC have shown rigid thinking and a degraded ability to perform new or novel behaviors due to perseveration and decreased abstract thinking (Goldberg, 2001; Fuster, 1989). Additionally, clinical patients with PFC impairment show increased distraction by irrelevant stimuli (Horne, 1993).

Research investigating the effects of sleep deprivation on PFC function has yielded results similar to that of patients with damage to the frontal lobes. The following sections describe the function of sleep in cerebral restitution as well as the physiological and cognitive effects of sleep deprivation on the PFC.

2.4 Sleep Psychophysiology

Sleep is typically quantified using polysomnography, in which brain electrical activity, eye movements, and muscle activity are measured simultaneously. For purposes of quantification, patterns of brain electrical activity, eye movements, and muscle activity have been artificially divided into five stages based on rules which were developed in 1968 (Rechtschaffen and Kales, 1968). Stage 1 is the lightest stage of sleep (in terms of arousability by auditory stimuli) and is usually the first stage humans enter during a normal, nocturnal sleep period. Stage 1 consists mostly of high amplitude, low frequency (4-8 cycles per second) brain electrical activity known as theta waves. Stage 1 is also characterized by slow, rolling eye movements and some decrease in muscle activity relative to the waking state. Stage 1 lasts between 5 and 10 minutes, during which time subjects transition to Stage 2. Stage 2 sleep is characterized by high-frequency (8-12 cps) bursts (generally 0.5 to 2 sec in duration) of brain

electrical activity known as sleep spindles. Stage 2 is also characterized by K-complexes, which are brain waves consisting of a sharp negative deflection followed by a sharp positive deflection (with a total amplitude of at least 75 microvolts) then a return to baseline, with each K-complex generally 0.5 to 1 sec in duration. Slow rolling eye movements may also be present in stage 2 sleep and muscle activity is generally below that seen during stage 1. Stage 2 gradually gives way to stages 3 and 4, which are the deepest stages of sleep (in terms of arousability) and are differentiated based on the amount of high-amplitude (at least 75 microvolts), slow-wave (0-3 cps) activity known as delta activity present (stage 3 is scored if at least 20% but no more than 50% of the 30-second scoring "epoch" consists of delta sleep; stage 4 is scored if 50% or more of the scoring epoch consists of delta sleep). Slow rolling eye movements may also be present in stages 3 and 4 but are often obscured by the brain electrical activity which can be picked up by the sensors used to record eye activity. Stages 3 and 4 are generally collectively referred to as slow-wave sleep (SWS).

Stages 1 through 4 are called "non-REM" (non-rapid eye movement) sleep. The fifth stage is rapid eye movement or "REM" sleep, so named due to the rapid eye movements (as opposed to the slow rolling eye movements of the non-REM stages) that characterize this stage. During REM, brain electrical activity is similar to that seen during stage 1 but is differentiated by the presence of "saw-toothed waves" which are similar to theta waves but have a distinctive saw-toothed top. REM is also characterized by muscle activity which is at its lowest level of the sleep period but not entirely absent. Short bursts of muscle activity also may occur during REM. Finally, the mentation associated with REM differs markedly from that of other sleep stages: REM mentation is characterized by emotions, storylike progressions, and bizarreness whereas non-REM mentation has been characterized as is characterized as more thought-like (Nielson, 2000). However, it should be kept in mind that characterization of REM and Non-REM mentation is based on the subject's recall of such mentation upon awakening. Thus, these characterizations may be a function of the post-awakening processes associated with each stage rather than an accurate reflection of the actual mentation.

A complete "sleep cycle" generally starts with stage 2 and ends with REM and lasts approximately 90 minutes. Humans generally pass through four to five such sleep cycles per night, with the relative distributions of stages 2, SWS, and REM changing with each sleep cycle (REM and stage 2 amounts increase and SWS amounts decrease). While the stages of sleep have

been well documented, the function of sleep with respect to cerebral restitution and sustainment of higher-order cognitive functions is less clear.

2.5 The Function of Sleep

Although the exact function of sleep is still unknown, there a general consensus that physical rest and sleep permit physical recuperation while sleep uniquely provides for cerebral restitution (Bennington, 2000). Kreuger, Obal, and Fang (1999) theorized that because the brain is constantly active throughout wakefulness, sleep permits brain activity to decrease, thereby allowing time for its tissues and chemicals to be restored. Furthermore, sleep may allow neurotransmitters to be "recharged" as a resupply of neurons is produced during sleep and not used again until wakefulness. In short, the exact recuperative nature of sleep (at least at the cellular level) remains unclear. However, in recent years, the neurophysiological (as measured by PET and fMRI) and cognitive (as measured by behavior) effects of sleep deprivation have been clarified.

2.6 Neurophysiological Effects of SD on the Prefrontal Cortex

With respect to the peripheral nervous system, sleep deprivation can produce mild activation of the HPA axis and also increase plasma concentrations of glucocorticoids (Meerlo, Koehl, van dar Borght, and Turek, 2002). Evidence shows that a primary physiological effect of sleep deprivation is a disruption in homeostasis and anabolic functions of the brain (Dinges, 2001). The effect of sleep deprivation on the immune system is unclear as few laboratory studies on the topic have been conducted; however, sleep loss may impair the immune system by lowering resistance to infection (Dinges, 1995; Dinges, 2001). Results of some studies have shown an association between sleep deprivation and heart disease (Partinen, 1994). In general, however, there does not appear to be a direct causal link between sleep deprivation and peripheral nervous system functioning. In contrast to peripheral functions, sleep deprivation appears to have a direct effect on cognitive function.

2.7 Cognitive Effects of sleep deprivation

Jones and Harrison (2001) summarized research to date showing sleep deprivation-induced decrements in cognitive functions associated with the frontal lobes. Several brain imaging studies have recently begun to look at the effects of total sleep deprivation on attention,

verbal fluency, and arithmetic tasks using fMRI and PET. In a study investigating the effects of sleep deprivation on cerebral function and behavioral responses, results indicated that additional brain regions were recruited to perform cognitive tasks to compensate for the effects of sleep deprivation on behavioral responses (Drummond and Brown, 2001). Thomas et al. (2000) investigated changes in regional brain activity and neurobehavioral impairments as a result of 24 hours of sleep deprivation. The results of PET data showed decreased activation of the entire brain, but specifically the PFC and thalamus. Decreased performance in serial addition and subtraction was associated with the deactivation of the prefrontal cortex, which is typically associated with working memory. In both of these studies, the extent to which brain activity resulted from sleep deprivation or task demands has not been determined.

In contrast to results showing decreased brain activation after sleep deprivation, the results of several studies have shown increased relative activation of additional brain areas after sleep deprivation. In a study investigating the effects of 35 hours of total sleep deprivation on regional brain activity during a divided attention task, results showed greater activation of the anterior cingulate, the right prefrontal cortex, and the parietal lobes after sleep deprivation than after a night of normal sleep. Interestingly, behavioral performance after sleep deprivation was not degraded (Drummond, Gillin, and Brown, 2001). The authors postulated that the increased activation of brain regions associated with attention indicated that additional resources were recruited to compensate for the sleep deprivation and allow the participants to maintain performance. Similarly, in a study on verbal learning after one night of sleep deprivation, Drummond et al. (2000) found that participants performed worse on a free recall test after sleep deprivation as compared to a rested state, but performed equally well on a recognition task. The fMRI showed the activation of the premotor area and temporal lobes during the task before sleep deprivation. After sleep deprivation, the temporal lobes were significantly more activated and, in addition, the bilateral parietal lobes and two frontal lobes were activated as well. Those areas that showed increased relative activation after sleep deprivation are associated with high working memory and cognitive loads.

The general conclusion formed from the mixed results of these brain imaging studies is that the brain's response to sleep deprivation is cognitive task-specific (Drummond et al., 1999). In other words, the brain may recruit additional resources to sustain performance on some

cognitive tasks and may show decreased cerebral response to others. None of the studies described above attempted to look at the brain's response to complex or higher-order tasks.

Physiological evidence suggests that sleep deprivation affects the function of the frontal lobes and specifically, the function of the prefrontal cortex. As the next section describes, sleep deprivation causes many cognitive decrements similar to those shown by patients with injury or damage to the frontal lobes. This provides further evidence of a link between sleep deprivation and the functions associated with the prefrontal cortex.

In general, sleep loss results in lapsing, cognitive slowing, memory impairment, and vigilance decrease (Himashree et al., 2002). The first behavioral signs of sleep deprivation include changes in mood and motivation, failure to complete routines, slower responses, physical exertion, and bickering (Giam, 1997). Simple task performance is degraded after both chronic and acute sleep deprivation including increased reaction time and decreased vigilance and attention. Moreover, higher order functions such as working memory, verbal fluency, logical reasoning, decision-making, and judgment are also degraded as a result of sleep deprivation (Thomas et al., 2000).

Evidence suggests that sleep deprivation has little effect on concrete, logical reasoning after 48 hours; however, decrements in innovative, flexible thinking and strategic planning have been shown after one night without sleep (as cited in Harrison and Horne, 1999). Horne (1988) found that participants showed increased perseveration and lack of flexibility after one night of sleep loss, even on tests lasting less than 10 minutes. Wimmer, Hoffmann, Bonato, and Moffitt (1992) attempted to duplicate the results of Horne (1988). Results indicated degraded flexible thinking and increased visual search time on a creative thinking and trail-making task, respectively. Working memory was not affected by sleep loss, nor was attention degraded. In further support of the research indicating that sleep loss decreases novel and innovative responses, Harrison and Horne (1998) found that sleep-deprived participants were slower and less successful at inhibiting strong word associations and producing unique or original word associations.

Several studies attempted to determine the effects of sleep loss on higher order cognitive functioning in a simulated real-world scenario. Bandaret, Stokes, Francesconi, Kowal, and Naitoh (1981) conducted an extended battle simulation during which military teams were forced to terminate the game after 45 hours sleep deprivation because they could not cope with

unexpected and swiftly changing task demands and were not able to update information based on new events. Another study on military personnel showed degraded innovative thinking, but sustained logical reasoning after two nights without sleep (May and Kline, 1987). Similar results have been found in the medical field showing doctors experienced degraded planning ability for unfamiliar tasks, but were still able to acquire and integrate large amounts of familiar information after one night of sleep loss (as cited in Harrison and Horne, 1999).

A study by Harrison and Horne (1999) investigated the effects of one night of sleep loss on innovative decision making using the complex business game Masterplanner. Masterplanner is a multistage, dynamic planning game in which both routine and novel responses are required to changing scenarios. Results indicated that sleep-deprived participants perseverated more than non sleep-deprived participants and showed fewer correct response shifts when the task required a change in reaction. The authors also tested the participants' ability to assimilate complex written material by administering the critical reasoning portion of the GMAT. Because scores on the GMAT were not degraded after one night of sleep loss, the authors postulated that participants were able to incorporate new information, but were not able to act upon it.

The previous sections have shown a physiological and behavioral link between the prefrontal cortex, higher level cognitive performance, and sleep deprivation. This link serves as a foundation for investigating the effects of sleep deprivation on executive function, complex task performance, and situation awareness. A physiological link between situation awareness has not yet been established using braining imaging techniques; however, if SA is considered as a higher level or executive cognitive process, achieving and maintaining situation awareness should be associated with PFC function. Before a link between sleep deprivation and impaired SA can be established, the SA construct must be better understood and operationally defined. The next section provides a brief overview of a human information processing model as a foundation for understanding and analyzing situation awareness.

2.8 A Model of Human Information Processing

Over the last fifty years, the psychology field has transitioned from modeling human behavior as stimulus-response relationships to modeling the cognitive processes responsible for transforming an input to an output. This new approach was termed information processing (Cooper, 2002). Within the context of human information processing, a distinction between

lower and higher level cognitive functions is often made. Although there is no clear distinction between higher and lower order cognitive function, Frith and Dolan (1996) emphasize that an important distinction between higher and lower order cognitive functioning is the subjective amount of mental effort required to perform a task. Lower-order cognitive functions are generally considered to be involuntary or automatic, and linked to individual brain areas. Higher order functions are considered to contain novel elements and to require controlled, strategic, or "executive" processes such as problem solving, critical thinking, concept formation, and reasoning (Frith and Dolan, 1996; Subbotsky). Executive processes coordinate multiple cognitive functions, allocate attentional resources, manage goals and priorities, and monitor the cognitive system as a whole (Cooper, 2002). It is generally thought that higher-order functions are based on lower-order functions such as attention, working memory, and temporal memory (Muzur, Pace-Schott, and Hobson, 2002). Despite the tendency for researchers to classify memory and attention as lower order functions, several studies indicate a link between higher order processes, working memory, and attention (as cited in Jones and Harrison, 2001). Thus, in complex tasks, the distinction between higher and lower functions may not be clear.

Models of human information processing provide insight into the cognitive processes underlying task performance. Figure 4 shows Wicken's (2000) model of human information processing stages.

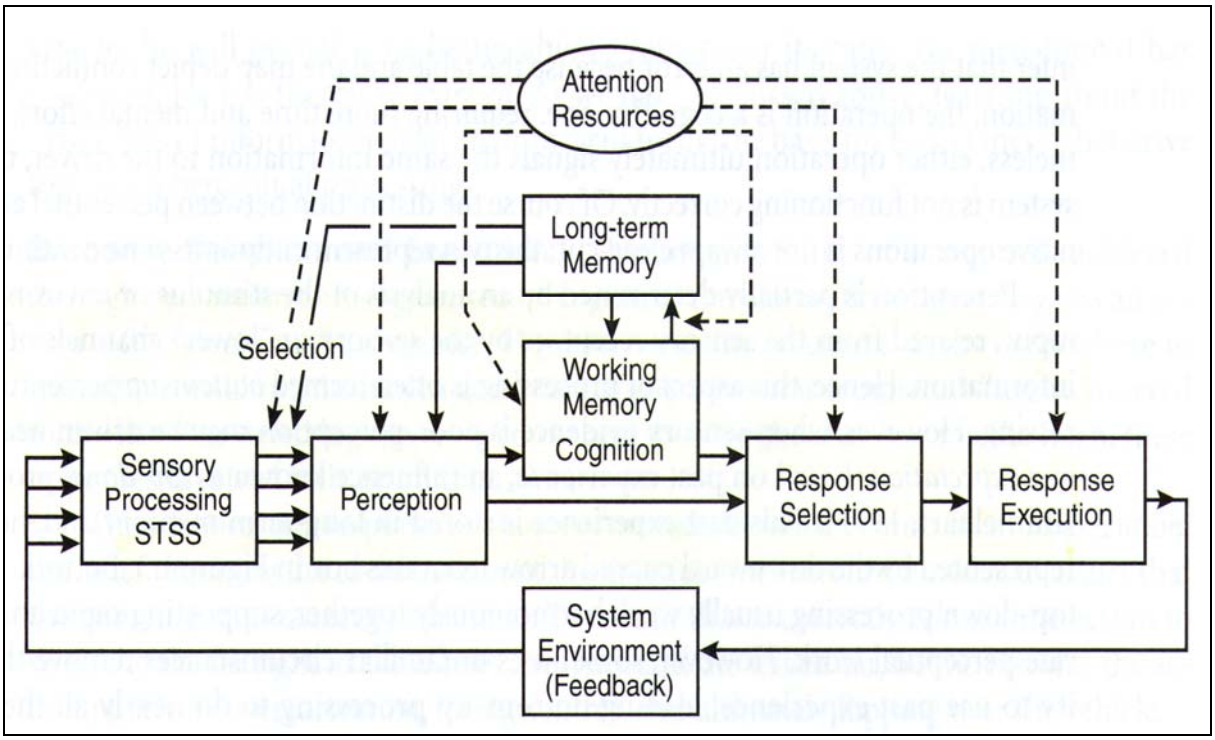


Figure 4. A model of human information processing (Wickens and Hollands, 2000)

In the sensory processing or short term sensory store (STSS) stage, the brain stores raw sensory information for brief periods of time, usually a few seconds. The sensory data is then relayed to the brain where it is perceived, or interpreted. Perceptual processing is generally an automatic and rapid process and is influenced by both sensory data and inputs from long term memory. Perception can be considered a bottom-up process when meaning is derived mostly from information in the sensory data. When perception relies primarily on inputs from long term memory to give meaning to sensory data, however, perception can be considered a top-down process. Based on the amount of sensory information available in a given input stimulus, the contribution of bottom-up and top-down processing will vary. Working memory is a temporary store of information used in the performance of cognitive operations. Cognitive operations are resource-limited conscious activities that manipulate, retain, or transform information. Working memory can transfer information into long term memory for storage through rehearsal. Long term memory influences working memory by providing expectancies for top-down processing. Response selection involves making a decision based on the cognitive transformations performed in working memory. Response execution is the controlled action of the selected response to accomplish a goal. Attention affects each stage in the human

information processing model by limiting the amount of mental resources available for each stage. And finally, feedback from actions is provided to the sensory system in an iterative manner (Wickens and Hollands, 2000). Wickens and Hollands (2000) extended their model to show the information processing components involved specifically in decision making (See Figure 5).

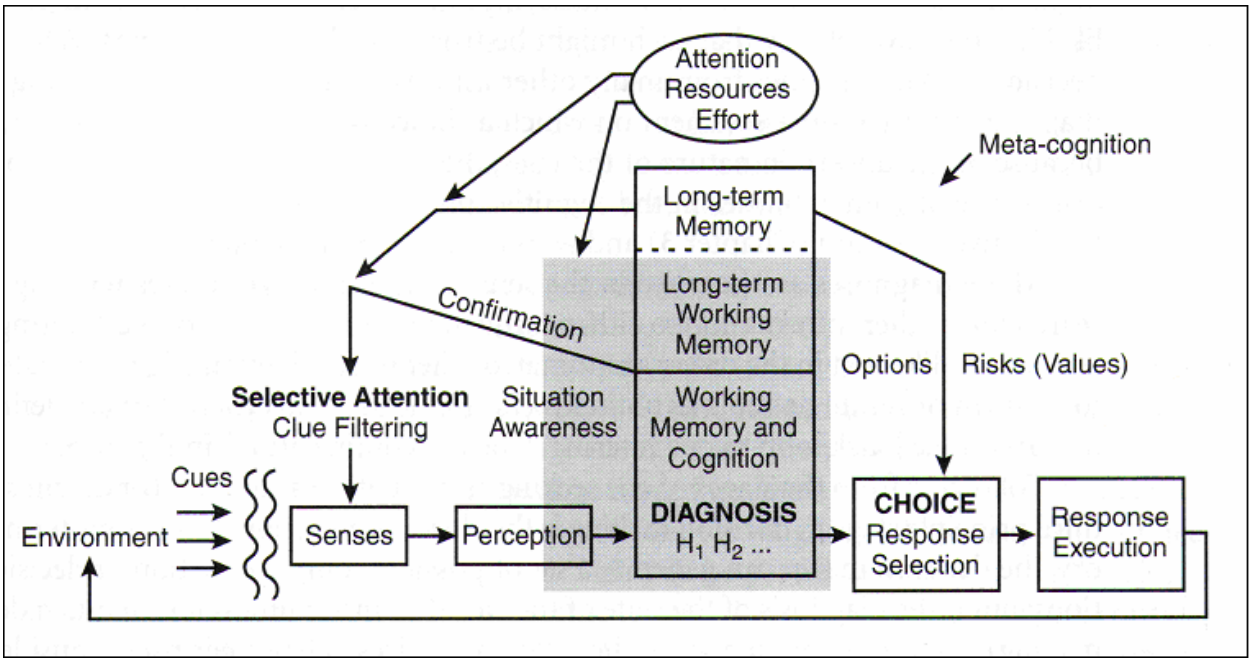


Figure 5. Decision making within the context of human information processing (Wickens and Hollands, 2000)

The decision maker seeks sensory cues from the environment and long term memory helps the decision maker selectively attend to particular cues based on past experience. The perceived cues form an awareness of the situation. In this diagram, situation awareness (SA) encompasses the combined operations of perception, working memory, and long term memory. SA then serves as the basis for hypotheses and expectancies about the system state. The decision making model of information processing is unique because SA may be incorrect due to uncertainty of the cues, incorrect mapping of cues to hypotheses, or impaired selective attention and working memory. The selection of action is influenced by long term memory which generates possible courses of action and determines the value and risk associated with each course of action. Feedback helps maintain and adjust situation awareness. The model also includes meta-cognition, or the awareness of one's knowledge, effort, and thinking which impacts the quality of decisions as a monitor of the overall decision making process.

Based on human information processing models such as the ones described above, researchers have attempted to define and model the construct of situation awareness. Despite a general understanding of the importance of SA, defining the construct has been more difficult. SA has been defined as a construct, a phenomenon, a process, and a state of knowledge and the debate over whether SA is a distinct phenomenon separate from decision making and action or a combination of cognitive subcomponents interacting during the operation of a dynamic, complex system continues (O'Hare, 1997). Notwithstanding the lack of consensus, several researchers have defined, explained, and modeled SA within the human information processing framework.

2.9 Defining Situation Awareness

The process of understanding and explaining SA has produced numerous definitions. Early definitions of SA included "a sixth sense," "mental representations of various flight relevant dimensions," an awareness of "his surroundings in light of his mission's goals" and "factors that will contribute to the safe flying of the aircraft under normal and non-normal conditions" (as cited in Carretta, Perry, and Ree, 1996; Sarter and Woods, 1991). Despite the sometimes abstract nature of earlier definitions of SA, common elements have included operator knowledge, understanding of goals, and tactical awareness (Selcon and Taylor, 1989). In the attempt to define SA, the aviation community focused on two main areas: the cognitive components of situation awareness and the temporal dimension (Sarter and Woods, 1991). Research identified cognitive components such as attention, perception, and working memory as contributors to SA (Endsley, 1999). The temporal domain of SA is also important because the specific cognitive components required at any given time during an SA-critical task vary depending on the task, environment, and tactical objective. A dynamic environment requires continuous perception and integration of information (Sarter and Woods, 1991).

A frequently cited definition of SA is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995a, p.36) Endsley described three levels of SA. In Level 1 SA, the operator must perceive the status, attributes, and dynamics of environmental elements that are relevant to tasks and goals. In Level 2 SA, the operator must integrate the perceived information and comprehend the significance of the combined information. And

finally, in Level 3 SA, the operator must be able to anticipate events (Endsley, 1999). The combined levels of SA are the basis for decision making and action of the operator.

Figure 6 depicts Endsley's (1995) model of SA in dynamic decision making. SA is the basis for decision making and resultant actions. Both SA and decision making are influenced by internal mechanisms such as goal-driven processing and information processing functions as well as outside factors such as stress, system design, and automation. Goal-driven processing directs the perception of information, allocation of attention, and the interpretation of information from the perspective of achieving a goal (Endsley, 1999). Thus, the perception, comprehension, and projection of the environment is influenced by the goals of the operator.

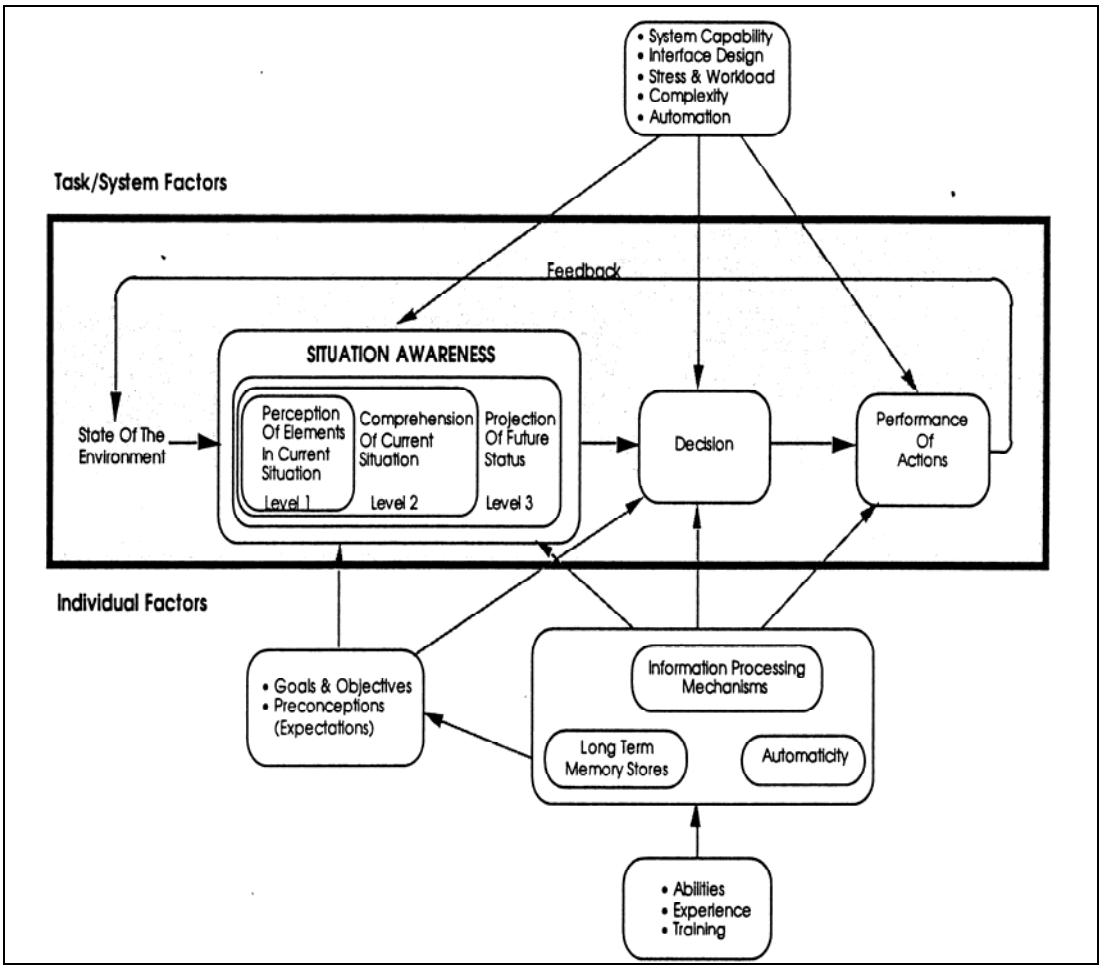


Figure 6. Situation Awareness in Dynamic Decision Making (Endsley, 1995b)

The information processing mechanisms affecting SA include short term sensory store, perception, working memory and long-term memory as the basic components upon which SA is formed. Level 1 SA requires preattentive processing to detect cues in the environment. For Level 2 SA, attention is needed for perceiving and processing information and divided attention is critical for operators to attend to multiple sources of information simultaneously. Poor Level 2 SA is often due to suboptimal sampling of information sources (Endsley, 1999). Moreover, projection of future status (Level 3 SA) requires the storage of present information and relevant rules for generating predictions of future states of the system. Working memory affects all levels of SA. Working memory can alter the perception of new information (Level 1 SA) through preconceptions or expectations about the information. Level 2 SA is facilitated by the integration of new information with previously acquired information within working memory to form an overall picture of the situation. Long term memory can facilitate SA through the storage of schemas and mental models that help interpret and integrate information. Schemas are frameworks for understanding information whereas mental models are explanations of system purpose and functioning (Endsley, 1995b). SA Levels 2 and 3 are considered higher order or executive processes.

Outside factors such as stress and system design also impact SA. Endsley (1999) asserted that some stress may improve performance by focusing attention to certain elements in the environment. However, high amounts of stress can degrade performance through attentional narrowing and reduced working memory capacity. System design can also facilitate or hinder SA through levels of complexity, automation, and workload. In summary, SA is attained through cognitive mechanisms (or subcomponents), but other factors influence the ability to achieve SA.

Endsley (2000b) emphasized that SA is a state of knowledge, separate from the processes used to achieve it (situation assessment). SA knowledge is derived from the integration of information from continuous situation assessments which, in turn, is a complex process requiring perception and pattern-matching, limited by working memory and attention capacity (Endsley, 1988). The interdependence of subcomponent processes and their effect on SA is important, however, the process of achieving SA is a separate from the state of SA. Different operators may use different cognitive processes to achieve the same knowledge about the situation and, conversely, operators may arrive at different states of knowledge about the situation using the

same cognitive processes. Providing information to an operator is not equivalent to providing an operator with SA (Endsley, 1999). The operator must perceive and attend to the stimuli in order to respond appropriately. In addition, operators may make good decisions despite having incomplete or inaccurate information and, conversely, operators with good SA may still make poor decisions (Endsley, 2000a). High levels of SA increase the probability of good decisions and low levels of SA decrease the probability of good decisions (Endsley, 2000b). She argues that decision making should not be a metric of SA because it is not necessarily a reflection of the level of SA, but rather a reflection of the operator's decision making ability. Understanding the processes underlying the state of SA only provides partial and indirect information about the operator's level of SA at a given time (Endsley, 2000b). Therefore, Endsley argues that outcome performance such as behavior and decisions are only indirect measurements of an operator's SA and should be considered separately (Endsley, 1990a).

In contrast to Endsley's view that SA is a state of knowledge, many researchers consider SA to be a process. Roscoe (1997) viewed SA as a unitary process requiring procedural, perceptual-motor, and decisional functions. The process of SA entails correctly perceiving information, allocating priorities, recognizing emerging opportunities, making decisions, ignoring distractions, tolerating frustration, and coping with boredom. Maintaining SA requires cognitive processing in the form of pursuit tracking, pattern recognition, spatial orientation, and short-term memory. Another process-related definition of SA was expressed by Hartman and Secrist (1991) who asserted that SA is principally cognitive information processing enriched by experience. They view SA as a phenomenon that facilitates and enhances individual cognitive skills and abilities that are essential to performance in a complex and dynamic environment. These skills and abilities were categorized as follows: sensory information acquisition, internal cognitive processing, decision processing, and motor response. SA is thus a skill exercised by all pilots which consists of cognitive abilities in the areas of information acquisition, information processing, and trend analysis and prediction. The authors assert that high levels of skill in these areas result in better SA performance. Smith and Hancock (1995) define SA as "adaptive, externally directed consciousness" (p. 138). The authors view SA as a dynamic process of directing consciousness at achieving a goal in a specific task environment and then generating behavior toward that goal. The result of the SA process is knowledge about the task environment and actions directed toward a goal. And finally, Hopkin (1993) views SA as a

higher-order attention management ability completely independent from individual cognitive abilities. Attention management includes selective attention and divided attention (time-sharing). Other components of cognitive information processing are not considered part of SA in this definition.

In support of the process approach to defining SA, several authors argue against considering SA as a state of knowledge separate from decision making and action. Albers (1999) states that the fundamental purpose of providing an operator with information is to improve SA such that a decision can be made. However, when the amount of information presented to an operator is increased, SA may increase, but excess information may delay time critical decision making. When presented with too much information, operators alter their decision making process from relying on experience to make judgments to relying on a detailed analysis of information. The term "analysis paralysis" describes the process of delaying decision making until it is too late because an operator is focused on the situation assessment rather than the decision (Marsh, 2000).

Endsley asserts that while good SA increases the probability of good decisions and poor SA decreases the probability of good decisions, decision making is not an accurate reflection of an operator's SA knowledge. In support of Endsley's assertion, Randel and Pugh (1996) found that SA is key determinant in decision making for experts, but not for intermediates and novices. In a simulation of electronic warfare, results showed that experts focused on assessing the situation correctly while novices focused on executing a correct course of action.

In contrast, Taylor (1990) argues that comprehensive measurement of SA requires an understanding of the active and dynamic nature of maintaining SA as well as the role of information in decision making. Furthermore, Wickens (2002b) acknowledges that an accurate understanding of the situation is necessary for performance; however, optimal action selection is not based solely upon SA. Other inputs to decision making such as internalized values and expectancies also play a role in performance. Endsley argues against using decision making metrics as a measure of SA because decision making is not necessarily a reflection of the operator's level of SA, but rather a reflection of the operator's decision making ability. However, that argument can be extended to say that metrics of SA knowledge are not a good predictor of outcome performance because they overlook the contribution of decision making to performance in a complex system. For example, Endsley (1990b) found mixed results when

using SA levels as a predictor of pilot performance. Results showed that high SA scores were correlated with high performance scores, but the converse was not true. Endsley attributed the discrepancy to pilots' ability to compensate for low SA by acting conservatively, but never addressed the role that decision making or other factors may have played in the pilots' performance outcomes. Therefore, while SA knowledge may be a necessary, it is not sufficient for operator performance.

Furthermore, an operational definition of SA cannot isolate the construct from the operational environment in which SA is required for performance in complex systems. In a statement by the Office of Naval Research, Marsh (2000) emphasized that situation awareness is not sufficient for the armed forces to dominate the battlespace. It is the ability to use knowledge and awareness to make decisions and take actions that is critical to the superiority of U.S. armed forces. Therefore, it is critical to understand the relationship between cognition, SA, decision making and action.

Depending on the author and the context, the definition of SA ranges from a process involving all cognitive processes to a state of knowledge independent from decision and action. Despite the various attempts to define SA, Flach (1995) cautions that while SA may be a real phenomenon, it is not necessarily an objective cause of accidents and injuries. Furthermore, he argues that attributing accidents to SA can be construed as circular logic. For example, stating that SA was lost because the operator responded incorrectly cannot be explained by saying that the human responded incorrectly because he lost SA. Sarter and Woods (1991) extend this argument to say that if SA is not an objective cause, SA may not be worthy of definition as a separable construct and that SA research is an example of "the tendency of applied cognitive psychology to coin new terminology in the face of ill-understood issues" (p.45). Thus, further research is needed to elucidate the SA construct and determine whether SA is, in fact, a separable construct.

Thus, this dissertation studied the relationship among executive function, decision making, and action during complex task performance in order to determine whether SA is indeed a construct separable from decision making and action or whether SA is better defined as a process involving all cognitive functions. Furthermore, this dissertation investigated the effect of sleep deprivation on the composition of the SA construct. A recent study by Boag (2003) found little empirical support for the three levels of SA as defined by Endsley. In a comparison of

three different SA measurement methodologies, Boag found that factor loadings were different depending on the SA measurement technique used (i.e., observation, subjective ratings, and objective measurement). To further explore Boag's findings, this dissertation attempted to determine whether the composition of the SA construct is dependent upon task requirements and environmental conditions (e.g., sleep deprivation).

2.9.1 An Operational Definition of SA

Based on conflicting literature on the SA construct, this dissertation proposed a process definition of SA adapted from Roscoe (1997). It was as follows: situation awareness is the process of perceiving information, appropriately shifting and sustaining attention, reasoning, allocating priorities, planning, making decisions, and coping with stressors during dynamic, complex task performance. Some researchers would argue that this definition does not distinguish SA from the cognitive information processing requirements of complex task performance and that individual components are not sufficient to define SA because there is some overarching component to SA that cannot be accounted for simply by looking at the individual parts. However, this dissertation proposed a hierarchy of information processing to better explain the relationship between lower-level cognitive function, executive function, and SA in which the levels of the hierarchy may not always be clearly distinguished because of the interrelationships among information processing components during complex task performance.

At the top of the hierarchy is complex task performance, herein considered to be the concurrent performance of multiple cognitive tasks (both lower and higher level) requiring effort. The next level in the hierarchy includes executive functions. Higher order functions are considered to contain novel elements and to require controlled, strategic, or "executive" processes such as problem solving, critical thinking, concept formation, and reasoning (Frith and Dolan, 1996; Subbotsky, no year). Executive processes coordinate multiple cognitive functions, allocate attentional resources, manage goals and priorities, and monitor the cognitive system as a whole (Cooper, 2002). It is generally thought that higher-order functions are based on lower-order functions such as attention, working memory, and temporal memory (Muzur et al., 2002). At the bottom of the hierarchy is lower level cognition, generally considered to be involuntary or automatic, and linked to individual brain areas (Subbotsky, no year; Frith and Dolan, 1996). This dissertation posited that situation awareness overlaps both the executive function level and the lower level of the hierarchy. It is not distinguishable from either lower level cognition or

executive function, but rather requires both depending on task demands. SA may be required for complex task performance, but is not sufficient by itself. Furthermore, SA may utilize executive processing, but executive processing does not always involve SA. Thus, SA is not separable from information processing in general and is considered to be only one component contributing to CTP.

There are both advantages and disadvantages to using multiple measures of cognitive ability to predict performance. Pilot and operator selection batteries have historically relied on tests of individual cognitive abilities to predict training performance and SA ability. Selection batteries have been used in both civilian and military aviation since World War I to reduce training costs, improve retention of pilots, and identify the characteristics of good pilots (Carretta and Ree, 2003). Test batteries typically relied on tests of individual cognitive ability such as reaction time, manual dexterity, memory, and spatial orientation, with little success (Damos, 1996). Numerous studies have shown the failure of test batteries to predict pilot performance. In general, test batteries have failed to account for more than 25% of the variance in performance and training success. Thus, administering a single test of cognitive ability is almost as predictive as administering the entire test batter (Roscoe, 1997). As an example, a study of 171 U.S. Air Force F-15 pilots compared performance on individual tests of working memory, velocity estimation, near threshold processing, reasoning, and spatial ability with observer ratings of SA. Results showed that variance attributed to single measures of cognitive ability was not predictive of SA (Carretta et al., 1996). The failure of test batteries to predict piloting skill or SA have been attributed to failure to demonstrate construct validity, lack of statistical power, failure to cross-validate studies, and misinterpretation of correlations and regression (Carretta and Ree, 2003). In addition, the tests have been validated on pass-fail criteria rather than specific criterion measures based on specific job requirements and therefore are too general to differentiate between levels of performance (Damos, 1996).

Despite its disadvantages, there are several advantages of using test batteries. Cognitive test batteries are easy to administer. Secondly, the tests provide direct measurement of subcomponent ability and the tests have been previously evaluated for construct validity. Finally, cognitive tests are sensitive to individual differences.

In contrast to traditional test batteries, this dissertation used a series of executive tasks (e.g., planning, decision making) to determine if complex task performance can be predicted by

executive function or if complex task performance did, in fact, require SA as a phenomenon above and beyond cognition for successful performance. Additionally, complex task performance was evaluated using specific performance metrics, not against pass-fail criterion (e.g., number of errors, completion time).

2.10 Tests of Executive Function

Tests of neuropsychological function were originally developed to differentiate brain-damaged individuals from control subjects. Beginning in the 1970's, researchers began to develop test batteries to measure specific cognitive function (Kolb and Whishaw, 1996). There are several approaches to neuropsychological testing. Standardized test batteries require little knowledge of the underlying theoretical basis for the test and have simple administration, scoring, and interpretation. Individualized assessments require substantial knowledge of the underlying psychological construct and rationale for the test so that qualitative assessments of cognitive performance can be made. Alternatively, standardized batteries can be given to each participant while using qualitative analysis to determine the cognitive deficits (Kolb and Whishaw, 1996).

Most sleep deprivation studies attempting to look at higher order functions have studied verbal fluency, logical reasoning, planning, and working memory (as cited in Muzur et al., 2002). Unfortunately, many tests of these "executive" functions have been shown to have poor construct validity and poor retest reliability. Testing the construct validity of the executive function tasks in this study was outside the scope of the dissertation, and thus, this dissertation used previously validated psychological tests with objective scoring methods to eliminate the need for qualitative assessments. In addition, the cognitive tests were selected based on their sensitivity to sleep loss and/or sensitivity to individual differences in PFC function, their hypothesized relationship to SA, and their face validity for requiring SA for good performance. The selected tests were as follows: Psychomotor Vigilance Task (sustained attention), Letter-Number Sequencing (working memory), Letter Sets (inductive reasoning), Logical Reasoning (deductive reasoning), Maze Tracing Test (planning), and Iowa Gambling Task (decision making).

2.10.1 Psychomotor Vigilance Task (Sustained Attention)

Sustained attention was measured using the psychomotor vigilance task (PVT). Sustained attention, or vigilance, has been shown sensitive to sleep loss in numerous sleep deprivation studies (Belenky et al., 1994; Belenky et al., 2003; Himashree et al., 2002). The psychomotor vigilance task was selected because of its demonstrated sensitivity to sleep loss and its lack of learning effects (Balkin et al., 2000).

2.10.2 Letter-Number Sequencing (Working Memory)

Working memory was tested using a letter-number sequencing (LNS) task. LNS is a subset of the Wechsler Adult Intelligence Scale working memory test battery. Reliability ranges from .70 to .93 (Titus et al., 2002). The LNS was shown to activate the frontal lobe in a PET study by (Haut, Kuwabara, Leach, and Arias, 2000). The PET scan found activation of the orbital frontal lobe, DLPFC, and posterior parietal cortex during performance of the LNS. Therefore, this dissertation used this test because it has been shown to activate the PFC.

2.10.3 Letter Sets and Logical Reasoning (Inductive and Deductive Reasoning)

Deductive and inductive reasoning were tested using the Logical Reasoning test and Letter Sets test, respectively. Deductive reasoning involves making specific inferences from general principles whereas inductive reasoning involves inferring general principles from specific facts (Martin, 2003). The reliability of the Letter Sets Test ranged from .74 to .84 and the reliability of the Logical Reasoning Test ranged from .46 to .64 (Ekstrom, French et al., 1976).

May and Kline (1987) conducted a field study to determine the effects of sleep deprivation on several cognitive processes during a sustained operations exercise. They found that performance on the Letters Sets test of inductive reasoning and the Logical Reasoning test of deductive reasoning (Ekstrom, French, and Harman, 1976) actually improved after two nights without sleep. These findings seem contradictory to the hypothesis that performance on tasks tapping the PFC should degrade after sleep loss. However, this study was conducted during a sustained operations field exercise in which the there was no experimental control over physical exertion, additional stressors, or the regulation of sleep. Therefore, this dissertation attempted to refute these findings under controlled laboratory conditions using the Letter Sets test and Logical Reasoning test.

2.10.4 Maze Tracing Test (Planning)

The Maze Tracing Test was derived as a test of spatial planning and has a reliability that ranges from .89 to .94 (Ekstrom et al., 1976). Although this specific maze tracing test has not been cited in literature regarding PFC-damage or sleep deprivation, a brain imaging study by Peterson et al. (1998) found that the right premotor and parietal cortex and left cerebellar hemisphere were activated during unpracticed maze tracing. Thus, this dissertation used the Maze Tracing Test as a measure of planning ability because a similar test was shown to activate the PFC.

2.10.5 Iowa Gambling Task (Decision making)

Bechara et al. (1994) originally developed the Iowa Gambling Task to measure the effects of prefrontal damage on a simulated real-life decision making task. The Iowa Gambling Task (IGT) has shown sensitivity to both frontal lobe damage and sleep deprivation. Results of the original Bechara (1994) study indicated that prefrontal patients performed worse on the IGT than controls and that the decrement was sustained for up to six months. The IGT was also used to assess decision making deficits based on the location of frontal lesions. Manes et al. (2002) found that patients with frontal damage (regardless of location) performed worse on the IGT than controls. More specifically, patients with dorsomedial damage selected more cards from risky decks. In addition, patients with DLPFC damage performed worse on the IGT than controls, but also showed additional deficits in working memory and attentional shifting. Research has also indicated ventromedial PFC and right DLPFC lesions result in degraded performance on the IGT (Bechara, Damasio, Tranel, and Anderson, 1998). IGT sensitivity to sleep deprivation was shown in a study by (Herscovitch et al., 1980). They found significant decrements in IGT performance after five nights of partial sleep deprivation. Based on evidence supporting the sensitivity of the IGT to PFC damage and partial sleep deprivation, this dissertation used this task to assess the effects of total sleep deprivation on decision making performance.

2.11 A Complex Task Hypothesized to Require SA for Good Performance

The Wondrous Original Method for Basic Airmanship Testing (WOMBAT) is a performance-based technique designed to predict SA ability in pilots, air traffic controllers, and nuclear power plant operators as well as to measure an operator's ability to maintain SA in a dynamic and complex environment (Roscoe, Corl, and LaRoche, 2001; *WOMBAT: Modern Psychological Assessment*). WOMBAT was selected as a complex, dynamic task for this

dissertation because the premise of WOMBAT is that complex operations such as situation awareness are not a collection of individually simple operations, but rather involve a unitary process requiring procedural, perceptual-motor, and decisional functions (Roscoe, 1997).

WOMBAT requires the operator to scan multiple information sources and integrate information about relevant events, conditions, and resources. The operator must quickly prioritize information and allocate attention accordingly (Roscoe, 1997). Situational complexity is built into the scenarios by combining individually simple subtasks such as those required for SA into an adaptive scenario that contains multiple sources of information and multiple response alternatives (Roscoe et al., 2001). The subcomponents in WOMBAT include the following:

- Identifying important information and allocating priorities accordingly
- Perceiving a situation correctly and being vigilant
- Discovering implicit rules through induction and deduction
- Recognizing emerging opportunities and acting upon them
- Ignoring distractions and tolerating frustration
- Coping with stress from high workload and poor performance indications
- Coping with boredom from route tasks and low workload

In order to encompass all of these demands and constraints on an operator, the WOMBAT tasks involve simultaneous performance of pursuit tracking, pattern recognition, spatial imagery, and short-term memory. The specific WOMBAT test scenario is described in detail below.

2.11.1 WOMBAT Methodology

WOMBAT contains four underlying tasks: target tracking, pattern recognition, spatial orientation, and short-term memory. The WOMBAT test score is comprised of points earned during the simultaneous performance of a primary tracking task and three secondary tasks. The goal is to score as many points as possible within the time allotted for the scenario, typically 90 minutes. The primary background task is a tracking task that requires the operator to monitor and search for targets on a grid using a left-hand, single-axis tracking task and a right-hand, dual-axis, tracking task. The combined performance of both tracking tasks comprises the Tracking Score and is displayed to the operator by the Performance-Worth indicator. The tracking task can be put into automatic mode (Autotrack) so that the operator can pursue other methods of scoring points. Autotrack is prone to failures over varying severity and, thus, must be monitored

by the operator during the performance of other tasks. Vigilance is required in that tracking performance must be maintained to avoid serious penalty.

The secondary task is known as the "bonus pouch" and is comprised of three tasks: 3-dimensional figure rotation and matching, quadrant location, and two-back serial digit canceling. The secondary tasks allow the operator to demonstrate situational awareness, short-term memory, temporal and spatial pattern recognition, and procedural compliance. Points and penalties for secondary tasks are assessed in various forms. Bonus points are indicated on the display. The operator may suspend bonus play at any time to attend to the primary tracking task without penalty. Time-sharing of attention is required to respond to autotracking failures while performing secondary tasks.

To perform well, the operator must monitor performance indicators and shift attention and priorities according to the potential scoring worths and current scoring rates. A task's high worth is maintained by performing that task often. In order to maintain high tracking worth, bonus tasks must be performed frequently. To maintain high bonus worths, the tracking task must be performed frequently. Moreover, every time a bonus task is performed, it loses worth and other tasks gain worth. Prioritization is critical because scoring on tracking and secondary tasks is determined by their momentary worth. Tolerance of frustration is tested when the computer produces a low rate of scoring (regardless of individual performance) at random intervals throughout the test scenario. The WOMBAT structure emphasizes a rational attention allocation strategy and high final scores are a result of effective management of task worths (Roscoe et al., 2001).

2.11.2 WOMBAT Validation

Few scientific studies investigating the validity of WOMBAT as an SA prediction tool exist. The limited number of studies performed are described here. O'Hare (1997) conducted two studies to determine whether WOMBAT test scores were related to individual component tasks similar to the ones tested by WOMBAT and to determine the validity of WOMBAT for predicting pilot performance. In the first study, demographic factors and performance on individual tests of component abilities were correlated to performance on WOMBAT. Tests were selected from the Walter Reed Assessment Battery to test pattern recognition, memory, visual search and recognition, and spatial ability because they were hypothesized to be similar to the tasks underlying WOMBAT. Results indicated that predictors of initial WOMBAT

performance included computer game experience, pattern recognition ability, and manikin performance. Results indicated that after 50 minutes of testing, only pattern recognition was a significant predictor of SA. After 60 minutes, however, no single component ability or level of computer game experience was a significant predictor of WOMBAT performance.

In the second study, performance on WOMBAT was correlated with pilot performance. The performance of eight elite pilots (as determined from superior performance in gliding competitions at national and international levels) were compared to other experienced pilots not classified as "elite" and to non-pilot controls matched for gender, age, and occupational status. Results showed that elite pilots performed significantly better on WOMBAT than non-pilot controls. Although the study indicates that high scores on WOMBAT are associated with elite piloting skills, it does not necessarily indicate that SA is the construct being measured. In an attempt to further validate WOMBAT, a 2000 study by O'Hare found that scores on WOMBAT predicted early performance on an air traffic control task that requires high SA. Moreover, the study attributed 72% of the variance in scores to SA (as cited in Roscoe, Corl, and LaRoche, 2001). Although there is preliminary evidence supporting the construct validity of WOMBAT with respect to SA, more controlled studies need to be conducted.

Aero Innovation, the distributor of WOMBAT-FC, advocates performing in-house predictive validation of WOMBAT by comparing WOMBAT test scores with objective performance measures on the system of interest (*Frequently Asked Questions*, 2003). The predictive validity of WOMBAT has been studied independently by flight schools such as the Centre Quebecois de Formation Aeronautique (CQFA), airlines such as Saudi Arabian Airlines, and other organizations which require complex human operations such as the Montreal Metro. CQFA evaluated WOMBAT as a pilot candidate selection tool between 1994 and 1996. Although they did not release the results of the analysis, CQFA gave a 70% weight to WOMBAT scores in the final candidate selection phase. As a result, the involuntary attrition rate during subsequent training fell to almost zero. Similarly, the Montreal Metro reported that after using WOMBAT as part of their train-traffic controller candidate selection process, attrition during training was close to zero and those candidates who scored better on WOMBAT received top marks as operational controllers. Unfortunately, data from this construct analyses was proprietary to avoid civil service issues (Roscoe, 2003).

A study by LeDuc et al. (1999) investigated the effects of sleep deprivation on WOMBAT performance. Although there were several methodological issues such as sample size and length of the WOMBAT test scenario, the study showed that WOMBAT was sensitive to sleep loss after 36 hours of sleep deprivation. Therefore, WOMBAT was used to assess complex task performance.

2.12 Research Needs

The preceding sections described recent attempts to determine the effects of sleep deprivation on several aspects of executive function including verbal fluency, temporal memory, logical reasoning, working memory, and planning as well as recent efforts to define and model the construct of situation awareness. To determine the factors that may contribute to accidents and injury in an operational setting, additional research is needed to validate anecdotal evidence about the causal nature of sleep deprivation in accidents and injuries as well as validate scientific data from previous studies investigating the effects of sleep deprivation on executive function. To date, few studies have attempted to determine the effects of sleep deprivation on a complex, interdependent task hypothesized to require SA for good performance. Therefore, as an initial step toward understanding the causal nature of accidents in an operational setting, research on the effects of sleep deprivation on executive function were expanded to investigate the effects of sleep deprivation on complex task performance over a lengthy period of time. In an operational setting, maintaining SA requires an ability to integrate information, to respond to novel situations, to think flexibly, to plan, and to make decisions with an appreciation for the potential consequences. All of these functions have been linked to the PFC and, as a result, all of these functions have the potential to be degraded by sleep deprivation. Before the effects of sleep deprivation on SA can be investigated, the SA construct must be better understood. This includes validating SA as a separable construct and investigating whether the composition of the SA construct changes as a result of sleep deprived conditions. Therefore, this dissertation investigated the effects of sleep deprivation on sustained attention and working memory as well as several executive functions including reasoning, planning, and decision making. In addition, his dissertation investigated the effects of sleep deprivation on SA as well as the construct validity of SA itself.

Chapter 3. Method

3.1 Participants

3.1.1 Subjects

The study included 48 total participants recruited from colleges and universities as well as Walter Reed Army Institute of Research (WRAIR). Age criteria ranged from 18-39 years of age. The age range has an upper limit of 39 years because previous research has indicated that other sleep parameters that may affect performance outcomes independent of sleep deprivation (e.g., total sleep time) change dramatically in middle-aged adults defined as 40 or more years of age (Wesensten et al., 2002).

3.1.2 Exclusion Criteria

Exclusion criteria were based on knowledge that a disease or condition is known to alter sleep and/or if the disease or condition may put the participant at additional risk (Wesensten et al., 2002). Participants were excluded if they had a history of or currently active condition of the following: HIV, hepatitis B or C (acute state), cardiovascular disease (to include mitral valve prolapse, cardiac enlargement or heart murmur (other than functional murmur), hepatomegaly, high blood pressure (to include resting blood pressure > 140/90 during the screening visit that does not decrease on a second reading taken at least 15 minutes later in the screening visit), asthma, renal disease, gastrointestinal disease, history of serious allergic reactions, immunological dysfunction, hematological disorders, cancer, endocrine or metabolic disorders, serious dermatologic disorders, adverse drug reactions, narrow angle glaucoma, or prostate enlargement. Participants were excluded if they had a clinically significant abnormal urinalysis, blood test, or echocardiogram, or if alcohol, nicotine, or drugs were present in their urine as determined by a urine drug screen. Participants were also potentially excluded if they had abnormalities in renal or liver function, history of epilepsy or any neurological disorder or damage, panic disorder, past or current use of licit or illicit psychoactive drugs, history of a sleep disorder, history of caffeine use in excess of 400mg per day on average, reported use of any drug which will not have been cleared from the body by 48 hours prior to participation as determined by a urine drug screen. Any participant with a history of treatment for psychological condition (e.g., attention deficit disorder, depression) or brain injury (e.g., concussion) was excluded from the study.

3.2 Measures of Executive Function

Tests of lower and higher order cognition were selected based on their sensitivity to sleep loss and/or damage to the prefrontal cortex and their hypothesized relationship to SA (as described in Chapter 2). The order of all cognitive tests was randomized to prevent an order effect. The tests analyzed the following cognitive functions: sustained attention, working memory, inductive and deductive reasoning, planning, and decision making. Descriptions of each task are found below.

3.2.1 Attention

Sustained attention was tested using the Psychomotor Vigilance Task (PVT). During this 5-minute test, participants monitored a hand-held computer display on which targets appeared at random intervals. Participants were instructed to press a button as quickly as possible when the target appeared on the screen (Wesensten, 2003). The estimated completion time for the PVT is five minutes. The dependent measure for this test was average reaction time.

3.2.2 Working Memory

Letter-number sequencing is a brief test of working memory, attention, and sequencing ability (Ryan and Paolo, 2001). Participants were presented orally with a series of letters and numbers in a non-sequential order (e.g., 3-N-7-B) and were required to maintain the items in working memory and reorder them numbers first, in order, followed by the letters, in alphabetical order (e.g., 3-7-B-N). Administration time was approximately 5 minutes. Equivalent forms were derived using a random letter-number generator. The dependent variable was total number of correct sequences.

3.2.4 Reasoning

Deductive and inductive reasoning were tested using the Logical Reasoning (Nonsense Syllogisms) test and Letter Sets test, respectively, selected from the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). The kit provided two parallel forms of each paper and pencil test for repeated testing.

The Logical Reasoning test was used to determine a participant's ability to reason deductively from premise to conclusion. The participant was presented with 15 nonsense syllogisms and required to determine whether the conclusions that follow from each premise were true or false. The time limit for test completion was four minutes. The dependent measure was the total number of correct responses. Incomplete responses will be counted as incorrect.

The Letter Sets test was designed to measure a participant's ability to form and test hypotheses that fit a set of data. There were 15 items containing five sets of four letters. For each item, the participant determined the rule that related four of the five sets together. Participants eliminated the set that did not follow the rule. The time limit was seven minutes to complete all 15 items. The dependent measure was total number of correct responses. Incomplete responses will be counted as incorrect.

3.2.5 Planning

The Maze Tracing Test was a test of planning selected from the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). Participants were given three minutes to trace a path through a series of mazes as quickly as possible without sacrificing accuracy. Participants were not penalized for lifting their pencil, retracing a path that led to a dead end, or for accidentally crossing lines at the sides of the path being taken. The kit provided two parallel forms of each paper and pencil test for repeated testing.

3.2.6 Decision Making

Decision making was tested using the Iowa Gambling Task. The participant was presented with four decks of cards from which to choose. Each time a card was selected, some amount of money was won. In some decks, money was also lost when the card was selected. Two of the decks were characterized by large wins with occasionally large losses. The other two decks were characterized by smaller wins, but also smaller losses. The goal was to win as much money as possible and lose as little money as possible. Administration time was approximately 20 minutes. The dependent measure was the total number of choices from risky decks (Manes et al., 2002).

3.3 Measures of Complex Task Performance

Performance on a complex task requiring SA was tested using WOMBAT-FC. The test consists of a primary air traffic control (ATC) task and three secondary tasks. The test required the participant to perceive information, allocate priorities based on new information, discover rules through induction and deduction, recognize emerging opportunities, ignore distractions, make decisions, and cope with boredom over a 60-minute test session.

WOMBAT contains four underlying tasks: target tracking, pattern recognition, spatial orientation, and short-term memory. The WOMBAT test score is comprised of points earned during the simultaneous performance of a primary tracking task and three secondary tasks. The goal is to score as many points as possible within the time allotted for the scenario. The primary background task is a tracking task that requires the participant to monitor multiple targets moving along pre-determined paths on a 5 x 5 grid. The participant can access information on a target's predetermined path at any time by clicking on it. Throughout the scenario, targets may collide. Although there is no penalty for allowing a collision to occur, extra points are earned for predicting and stopping a collision. The tracking task can be put into automatic mode (Autotrack) so that the operator can pursue other methods of scoring points. However, Autotrack is prone to failures over varying severity and, thus, must be monitored by the operator during the performance of other tasks. When the Autotrack fails, a target will "go missing." The target continues along its path; however, it is no longer visible on the screen. The participant must notice that a target is missing, remember where it should be along its path, and return the computer to Autotrack mode. The number of targets on the screen at one time is determined by how well the participant keeps track of existing targets. As the participant becomes more skilled, more targets will appear.

The secondary task is known as the "bonus pouch" and is comprised of three tasks: 3-dimensional figure rotation and matching, quadrant location, and two-back serial digit canceling. The secondary tasks allow the operator to demonstrate situational awareness, short-term memory, temporal and spatial pattern recognition, and procedural compliance. Points and penalties for secondary tasks are assessed in various forms. Bonus points are indicated on the display. The operator may suspend bonus play at any time to attend to the primary tracking task without penalty. Time-sharing of attention is required to respond to autotracking failures while performing secondary tasks.

The night prior to baseline testing, participants were given 60 minutes to read through the computerized instructions and ask questions. Participants were then given a 60-minute practice test session. The actual WOMBAT testing session lasted 60 minutes. The dependent variable was Overall Score. Additionally, performance scores were calculated automatically by WOMBAT at five-minute intervals.

3.4 Task Scheduling

The complete schedule of study events can be found in Appendix X. Table 3 shows the schedule of cognitive testing. The PVT was administered every two hours. Tests of executive function were administered in 5.5 hour testing blocks. Tests were given in the same order for each session. To prevent an order effect, the equivalent test forms were counterbalanced. With the exception of the IGT, all cognitive tests were administered sequentially. The sleep deprived testing block began at approximately 23 hours of sleep deprivation and ended at approximately 28 hours of sleep deprivation.

Table 3. Cognitive Testing Schedule

Baseline				**Sleep Deprived**			
Day	**Time**	**Test**	**Sleep Loss (hrs)**	**Day**	**Time**	**Test**	**Sleep Loss (hrs)**
THU	1150	IGT	4.8	FRI	0550	IGT	22.8
THU	1520	WOMBAT	8.3	FRI	0920	WOMBAT	26.3
THU	1620	PVT (5)	9.3	FRI	1020	PVT (14)	27.3
THU	1630	MTT	9.5	FRI	1030	MTT	27.5
THU	1635	LNS	9.6	FRI	1035	LNS	27.6
THU	1640	LS	9.7	FRI	1040	LS	27.7
THU	1650	NS	9.8	FRI	1050	NS	27.8

3.5 Measures of Objective Sleep/Wakefulness/Alertness

3.5.1 Polysomnographic (PSG) measurements

PSG measurements were recorded to identify sleep and wakefulness during the following the following periods of the study: during baseline sleep, during Maintenance of Wakefulness Tests (MWT), during WOMBAT testing, and during recovery sleep. PSG measurements included electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (EKG). Signals were recorded continuously using the Compumedics Siesta System. Contralateral mastoid leads served as references for all unipolar measurements (EEG and EOG). Electrodes were attached to the surface of the skin with surgical tape or collodion. PSG records were for 30-second epochs. Each epoch was assigned a stage, consisting of wake, and sleep states 1, 2, SWS, or REM. Sleep/wake scoring was conducted by experienced sleep

scorers whose scoring reliability meets or exceeds 85% agreement with scoring conducted by a board-certified clinical polysomnographer.

3.5.2 Repeated Tests of Sustained Wakefulness (RTSW)

During this test, participants were escorted to their individual darkened, sound-attenuated bedroom and told to lie down in their beds, close their eyes, and try to stay awake. PSG was not monitored online because the staff was not qualified to determine when the first 30-second epoch of unambiguous stage 2 sleep occurred (the criteria for test termination). Therefore, participants were wakened after 15 minutes and data was analyzed post-hoc for sleep latency.

3.6 Other Measures

3.6.1 Vital Signs

Tympanic temperature, blood pressure, and pulse were recorded upon arrival on Day 1 and throughout the study at regular intervals to monitor general health status by study technicians. Tympanic temperature was taken using a tympanic thermometer. Blood pressure and pulse were taken using an automated blood pressure and pulse monitor.

3.6.2 Urine drug screen

Urine was collected for drug screening upon arrival on Day 1 and every 24 hours for the remainder of the study. Samples were sent to a commercial laboratory to test for amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine metabolite, ethyl alcohol, methadone, nicotine metabolite, opiates, phencyclidine, phenothiazines, propoxyphene, and tricyclic antidepressants. No participants were dismissed from the study due to positive drug screens.

3.7 Equipment

The PVT was administered using a handheld Palm computer. The IGT was administered using a PC. The WOMBAT system was installed on a PC and the WOMBAT console (consisting of two joysticks and a keypad) was placed on a desk in front of the PC. The remaining tests were administered using pencil and paper versions.

3.8 Procedure

3.8.1 General

The study was run with four participants per session. The study lasted 65 hours including arrival, briefing, training, testing, and recovery sleep. Participants were sleep

deprived for a total of 40 hours; however, the sleep-deprived testing session ended at 28 hours of sleep deprivation. Previous studies have shown significant cognitive effects of sleep deprivation on vigilance and executive function in a range of 24 to 36 hours (Belenky et al., 1994; Harrison and Horne, 1998, 1999; Wimmer et al., 1992).

Laboratory facilities included two sleep suites. Each suite contained four bedrooms (one for each participant), a common area with sofas and a television, and a kitchen. Bathrooms are located directly outside the suites. At any time when testing was not being conducted, participants were allowed to pursue normal activities of daily living within the laboratory. These activities included watching television, reading, or playing video games. Physical exercise was not permitted.

There were no formal restrictions on food and water intake with the exception of any food or beverage containing caffeine or other stimulants. All meals were provided and participants were not allowed to bring their own food, drink, or dietary supplements.

One week prior to the study, participants wore an actigraph to monitor activity. In addition, participants completed a sleep log to record daily sleep habits. At 1800 hours Day 1 (Wednesday), subjects reported to the laboratory. They were oriented to the sleep suites and briefed on the schedule of allowed activities and restrictions. Vital signs were taken and a urine sample collected for drug analysis in all volunteers and for pregnancy screening in women. Polysomnographic recording electrodes (EOG, EKG, EMG, C_3 and C_4 EEG sites) were applied. Volunteers were also given instructions and familiarized with the PVT. Participants were given up to 60 minutes to read through the computerized WOMBAT instructions followed by a 60-minute practice test session. At approximately 2300, each volunteer was escorted to his/her own comfortable, sound-attenuated bedroom where they were allowed to sleep undisturbed from 2315 to 0700 Day 2 (Thursday). Upon awakening, vital signs were taken and volunteers were allowed to eat a meal. Baseline cognitive testing began at 0800 Day 2 with the PVT. All volunteers were administered a brief medical examination by the study physician prior to being cleared for release from the study. During the medical exam vital signs were taken (pulse, blood pressure, and temperature) and the volunteer was asked about any symptoms or complaints he or she might have had. If the study physician found the volunteer in good health and there were no significant symptoms or complaints, the volunteer was cleared for release from the study. Volunteers were then debriefed and released.

3.8.2 Compensation

Participants were paid $1600 for their participation in the study.

3.9 Forms

Participants completed several forms during a screening visit to determine eligibility for inclusion in the study. The Beck depression inventory was administered to screen for potential clinical depression. A State-Trait anxiety inventory was given to screen for potential clinical levels of anxiety. A preliminary sleep questionnaire was administered to screen for potential sleep disorders or sleep patterns outside the normal range. A morningness-eveningness questionnaire was given to determine whether they are morning-preferring or evening-preferring. A medical history form was completed by the participant to screen for potential current and past medical problems. A medical evaluation was then completed by a study physician.

During the study, participants filled out a medical health update to assess general health status and determine if participants followed pre-study procedures. An exit physical examination form was completed to determine if participants could be cleared for release from the study.

Chapter 4. Results and Discussion

The purpose of this dissertation was to form a better understanding of the effects of sleep deprivation (SD) on real-world, complex task performance requiring SA. To that end, the problem was approached by first investigating the effects of sleep deprivation on lower-level cognitive processes (e.g., attention and working memory). This information was then combined with analyses on higher order effects to (a) better understand the connection between lower and higher order effects of SD; (b) predict performance on a complex task (CTP); and (c) to form hypotheses about higher-order constructs such as SA. Ultimately, paired t-test results from the present study indicated there was no degradation of higher order cognitive function as a result of sleep deprivation; however, the results did provide insight into the relationship between lower-order cognition, executive function, SA and complex task performance. Additionally, a comparison of SA data collected under baseline versus sleep-deprived conditions provided insight into the complexity of understanding the SA construct as well as the potential task and situation-dependent nature of the construct. This dissertation laid some groundwork for understanding the interaction between lower and higher-order cognitive effects of SD and, at the same time, took a novel approach to investigating the SA construct by comparing the baseline vs. sleep deprived SA structure. Perhaps most importantly, results of this study suggest areas of future research necessary to reach the ultimate goal of understanding and predicting the effects of sleep deprivation on complex task performance.

This chapter consists of both results and discussion. First, a brief description of study sample demographics is presented. Next, the effects of sleep deprivation on subjective sleepiness and lower order cognitive function (e.g., attention, working memory) are discussed, followed by the effects of sleep deprivation on executive function and the link between the results for lower vs. higher-order cognition. In the third section, several explanations for the results attained in this study are suggested, including individual differences, cognitive processing hypotheses, and procedural and analytical limitations of this study. The composition and validity of the SA construct is discussed within the context of analyses on the effects of sleep deprivation on executive function and SA. Engineering recommendations formed specifically from this study are then proposed. In addition, more general future research needs are identified based on

the limitations of this study as well as gaps in the literature. Finally, the complexity of reaching the ultimate goal of understanding and sustaining CTP under sleep-deprived conditions is discussed.

4.1 Demographics

The study sample consisted of 48 total participants (26 M, 22 F). Figure 7 shows the age distribution of study sample. Participants' ages ranged from 18 to 33 (M=22.7, SD=3.6) and were normally distributed (p>.15) (see Appendix B). This distribution is similar to those found in previous sleep deprivation studies conducted at the Walter Reed facility (unpublished observations).

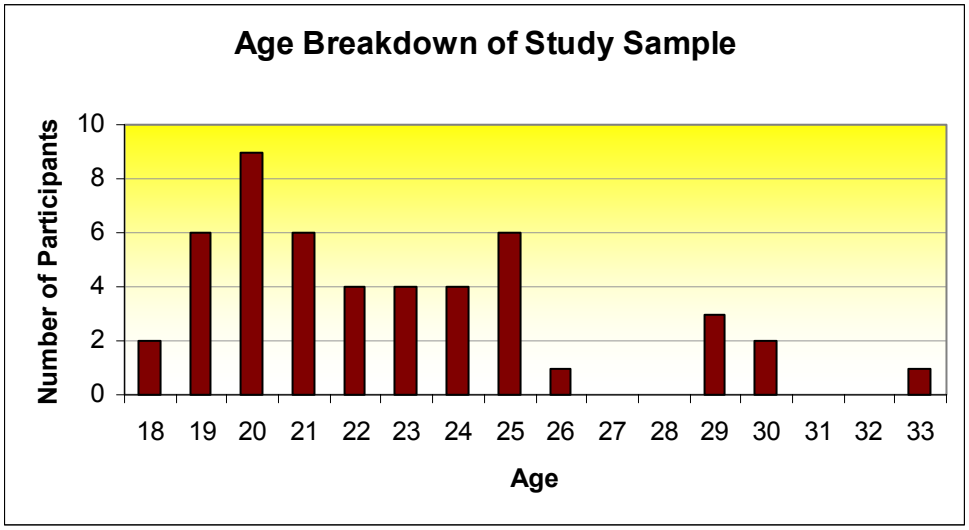


Figure 7. Age Distribution of Study Sample

Figure 8 illustrates ethnic distribution of the study sample. Ethnicity was based on participants' response to the question, "For demographic purposes, what would you like your race listed as?" Participants responded as follows: 15 indicated white/Caucasian ethnicity; 23 indicated African American ethnicity; 3 indicated Hispanic/white ethnicity; 1 indicated Hispanic/African American ethnicity; and 6 indicated Asian/Pacific Islander ethnicity. The ethnic distribution of the study sample was not representative of the military population. contained a larger percentage of African American participants than the military population and a smaller percentage of white participants than the military population (see Appendix B).

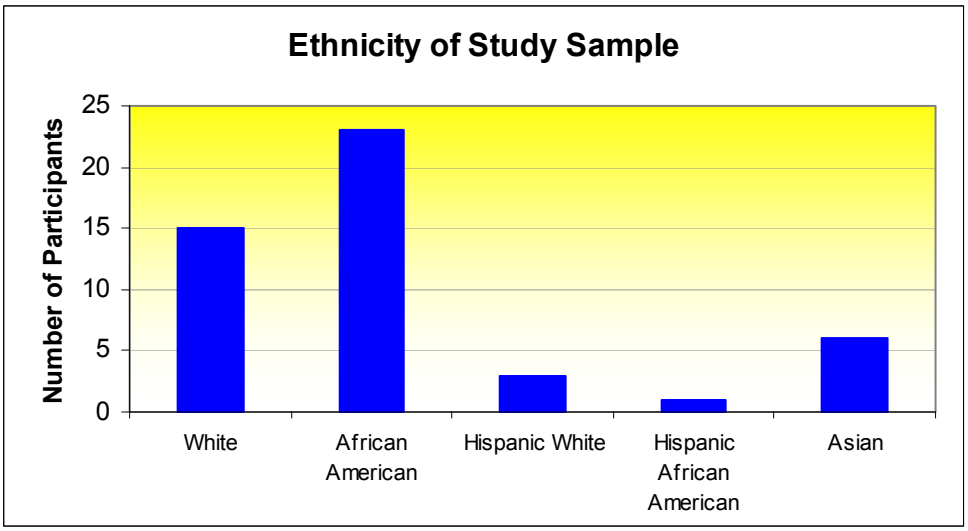


Figure 8. Ethnicity of Study Sample

The occupations of participants (as indicated by the participants on their medical history form) were broken down into four categories as follows: student, active duty military, unemployed, and other/not specified. Figure 9 shows the distribution of occupations. Twenty-five participants indicated that they were students; 7 indicated that they were active duty military; 2 indicated that they were unemployed; and 14 indicated that their occupation was "other" or they did not specify an occupation (specifying occupation was voluntary).

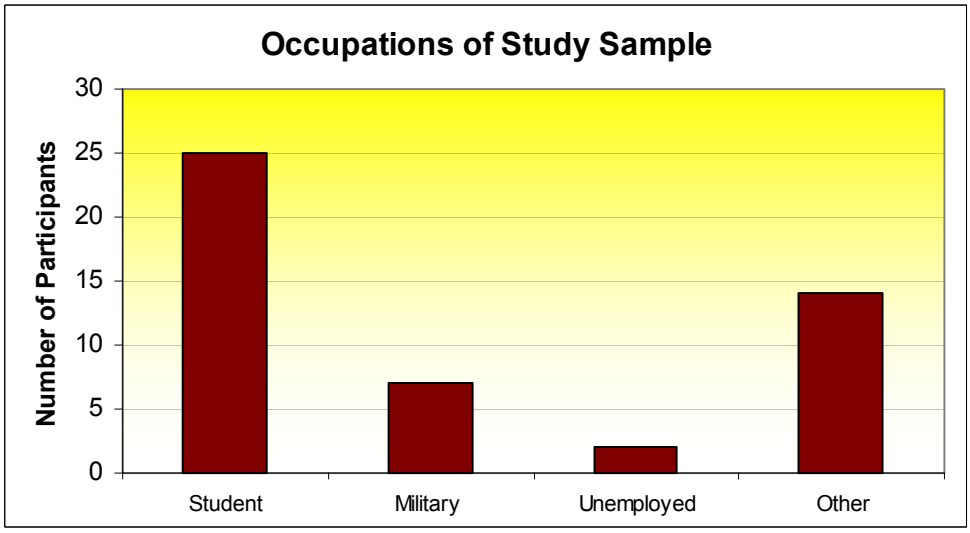


Figure 9. Occupations of Study Sample

Participants were asked prior to testing to describe their language background as one of the following: (1) have spoken only English since age three or younger; (2) have spoken English as well as another foreign language at home since age three or younger, and (3) learned English as a second language after age three. Figure 10 shows the language background of the study participants. Those participants who learned English after age three were classified as English as a second language (ESL) because research has shown that the fundamental structures of language are acquired within the first three years of life (Vihman and McLaughlin, 1982) and researchers consider simultaneous bilingualism to be distinct from sequential bilingualism (Langdon and Merino, 1992).

Thirty-five participants responded that they spoke only English; 5 participants responded that they spoke both English and another language from age three or younger; and 8 participants responded that they had learned to speak English after age three.

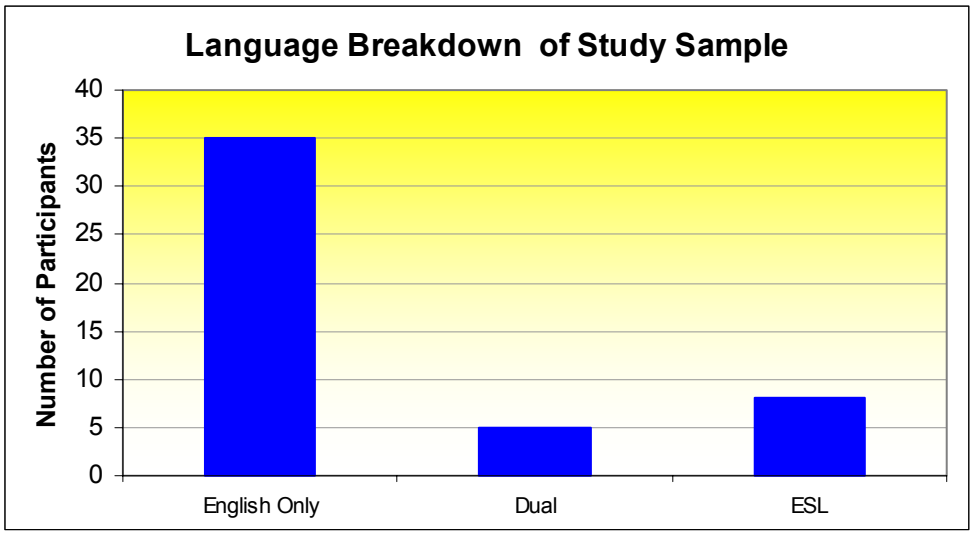


Figure 10. Languages Background of Study Sample

Figure 11 shows the number of males and females that participated in the study. Twenty-six males and 22 females participated.

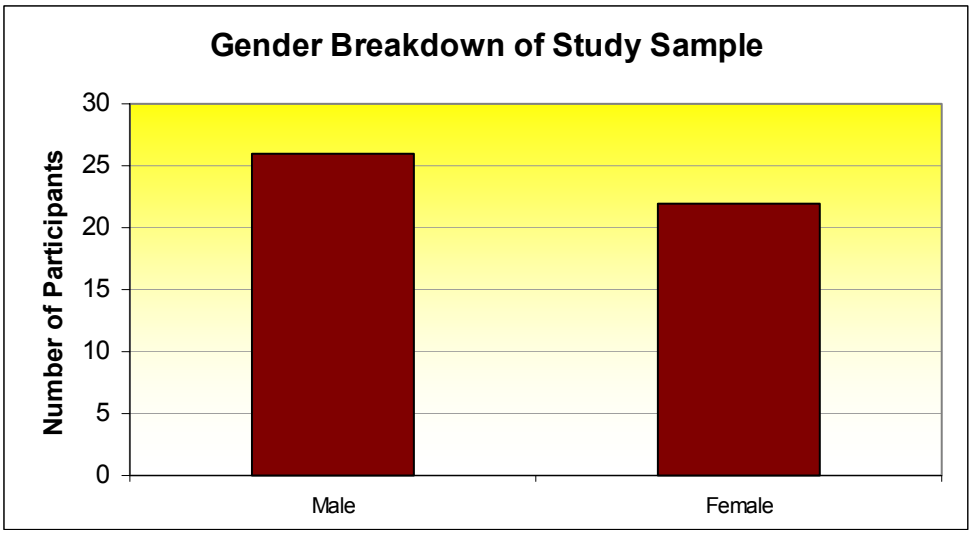


Figure 11. Gender of Study Sample

Figure 12 shows the distribution of self-reported education level of participants. Self-reported education ranged from 12 to 18 years (M=14.5, SD=1.8) and was normally distributed (p>.15) (see Appendix B).

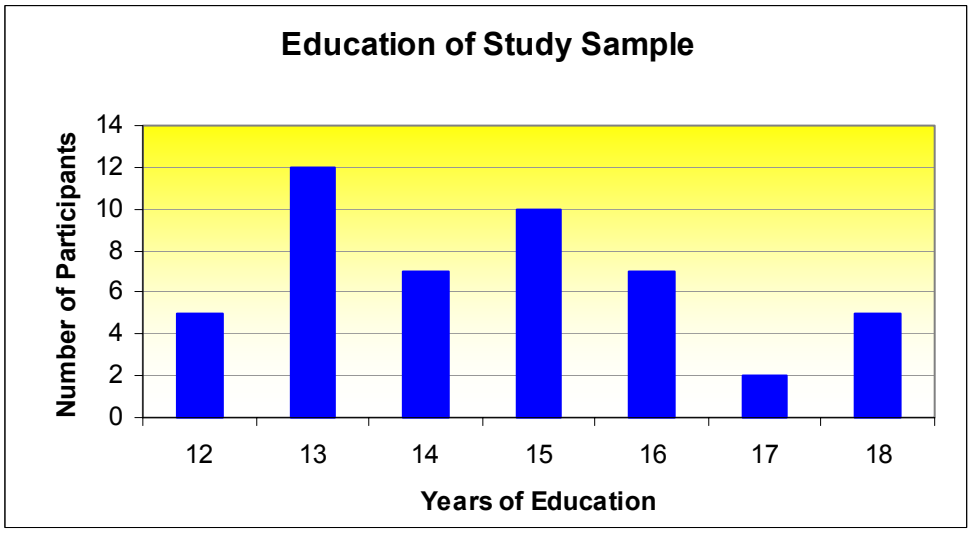


Figure 12. Education of Study Sample

Figure 13 shows the distribution of morningness-eveningness scores from the Morningness-Eveningness Questionnaire (Horne and Ostberg, 1976). For reference, an MEQ score of 41 or below is indicative of an "evening type;" scores ranging from 42 to 58 are

considered indicative of "intermediate types;" and an MEQ score 59 or above is indicative of a "morning type" (Horne and Ostberg, 1976). In the present sample, MEQ scores ranged from 26 to 70 (M=51.4, SD=9.6) and was normally distributed (p>.15) (see Appendix B). On a separate questionnaire, participants were also asked to subjectively classify themselves as a "morning type," an "evening type" or "both types." On this latter questionnaire, 20 participants described themselves as "morning types;" 15 described themselves as "evening types;" and 13 described themselves as "both morning and evening types." A significant correlation was found between MEQ score and self-classification (r=.372, p=.009). Figure 14 shows the comparison of self-reported morningness-eveningness and morningness-eveningness as determined from MEQ score. For purposes of demographic analysis, participants were grouped according to the Horne and Ostberg (1976) classification strata.

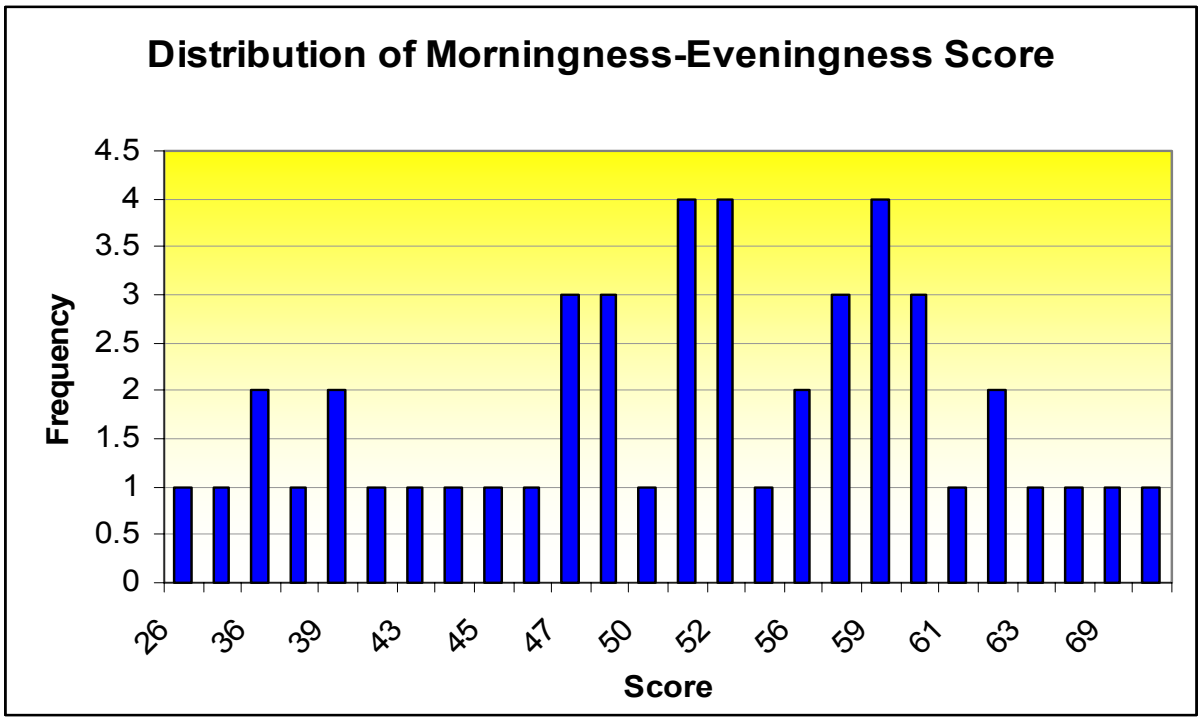


Figure 13. Histogram of Morningness-Eveningness Scores

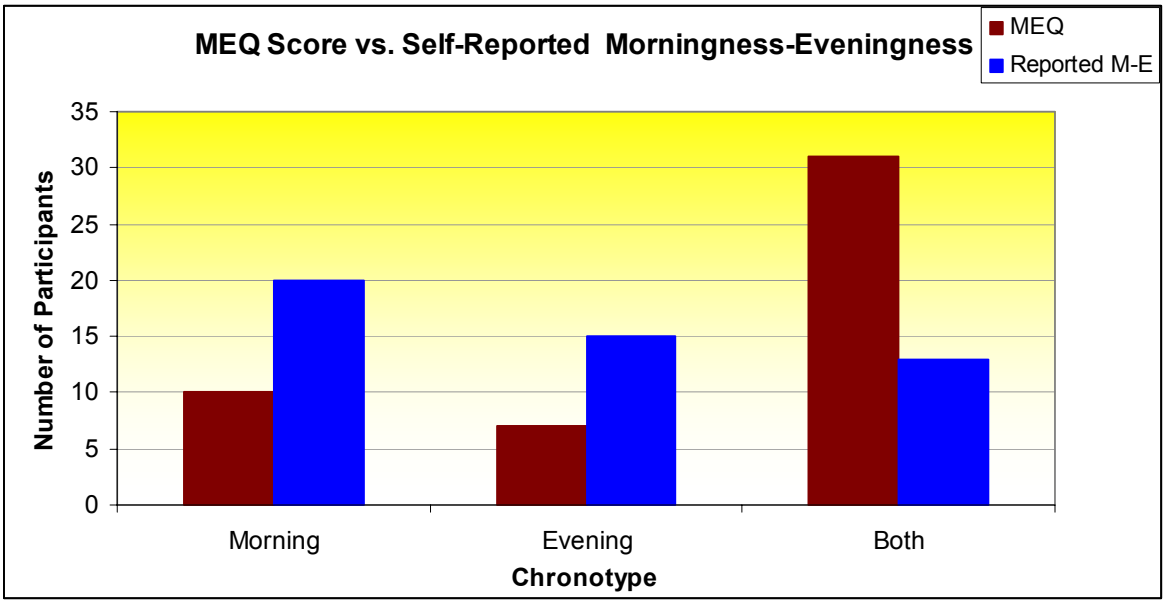


Figure 14. Self-reported vs. MEQ Morningness-Eveningness

In addition to the above demographic data, participants were evaluated using the Wechsler Abbreviated Scale of Intelligence (WASI), a short from of the Wechsler Adult Intelligence Scale—III (WAIS-III). The WAIS-III is considered the gold standard of intelligence testing in the United States (Wymer, Rayls, and Wagner, 2003) and has been used to assess learning disabilities, attention deficit/hyperactivity disorder and other neurological conditions (Wymer, Rayls, and Wagner, 2003). The WASI consists of the following subtests taken from the WAIS-III battery: Vocabulary, Similarities, Block Design, and Matrix Reasoning. Three scores can be calculated from these four tests: Verbal IQ (Vocabulary + Similarities), Performance IQ (Bock Design + Matrix Reasoning), and a Full 4 IQ score derived from all four tests. Because the WASI Full 4 IQ score correlates highly with the WAIS-III score (r=.82; Axelrod, 2002) it was selected for use in the present study. The WASI Full 4 IQ score was included as a demographic measure for two reasons: (1) as a potential means of explaining individual responsivity to sleep deprivation (i.e., a potential covariate for data analyses) and (2) as a potential means of assessing the validity of WOMBAT as a complex task rather than a test battery.

Figure 15 shows the distribution of Wechsler Abbreviated Scale of Intelligence (WASI) Full 4 IQ scores. The scores ranged from 84 to 129 (M=106.7, SD=12.1) and were normally distributed (p>.15) (Appendix B). Five participants scored in the 80-89 range, 25 participants

scored in the 90-109 range, 9 participants scored in the 110-119 range; and 9 participants scored in the 120-129 range. No participant scored in the 130+ range.

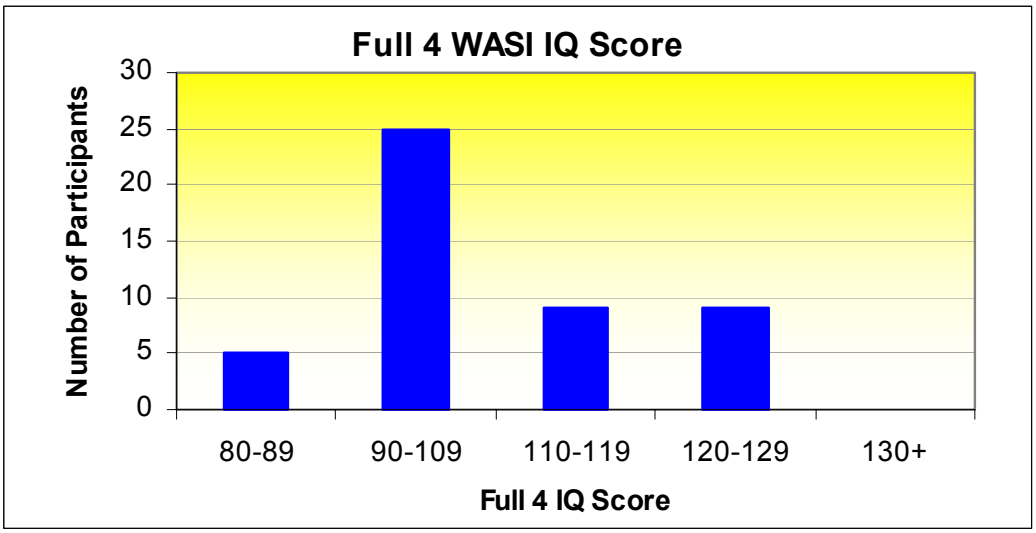


Figure 15. WASI Full 4 IQ Scores

Table 4 lists a comparison of the study sample distribution of WASI scores with normed WASI scores (N=2,245) (Wechsler, 1999). No participants fell into the "extremely low or borderline" categories. Fifty-two percent of participants fell within the "average" range, 19% fell in the "high average" range, and 19% fell in the "superior" range. No participants fell in the "very superior" range. The study sample distribution was slightly skewed (skewness=0.05) toward higher scores.

Table 4. Score Ranges and Descriptions

		Percent Included	
IQ Score	Classification	Theoretical Normal Curve	Actual Sample
130+	Very superior	2.2	0.0
120-129	Superior	6.7	19.0
110-119	High average	16.1	19.0
90-109	Average	50.0	52.0
80-89	Low average	16.1	10.0
70-79	Borderline	6.7	0.0
69 and below	Extremely low	2.2	0.0

One-way ANOVAs with Tukey HSD correction for multiple comparisons were conducted to determine if there were significant differences in WASI score based on

demographic variables (See Appendix B). Significant differences in WASI IQ scores were found between groups for the following demographics: education (p=.002), occupation (p=.000), and ethnicity (p=.000). Participants with 12 years of education scored significantly lower than those with 14 (p=.010) or 15 (p=.002) years of education. Participant with 15 years of education scored significantly higher than those with 16 years of education (p=.031). Education was still significant (p=.007) when age was used as a covariate. Student participants scored significantly higher on the WASI than active duty military (p=.001) and "other" occupations (p=.023). White participants scored significantly higher than African American participants (p=.000), Hispanic-African American and Hispanic white participants (p=.000), and Asian/Pacific Islander participants (p=.006). When age and education were used as covariates, there was still a significant difference in IQ score based on ethnicity (p=.000).

Despite the widespread use of intelligence testing in clinical, school, and research settings, there is an ongoing controversy over the cultural bias inherent in IQ testing. Research has focused on whether intelligence test scores are influenced by the environment and whether scores reflect acculturation into affluent white society rather than cognitive ability (Vincent, 1991). The results from this study indicated a difference in WASI scores based on ethnicity even when age and education were used as covariates. Therefore, caution was taken when interpreting and generalizing WASI-related findings to the population as a whole.

The demographic data presented in this section were used in subsequent analyses to investigate whether demographic characteristics of study participants were related to resilience to sleep deprivation. In the next section, the effects of sleep deprivation on subjective sleepiness and the relationship between sleepiness and attention (as measured by the PVT) are discussed.

4.2 Self-Reported Sleepiness, Morningness-Eveningness, and Demographic Characteristics

A Pearson correlation analysis ($\alpha = .05$) was performed to determine if MEQ score, self-reported morningness-eveningness, age and IQ score were significantly correlated with subjective sleepiness ratings on both Karolinska and Stanford sleepiness scales. Table 5 shows correlations for both baseline and sleep deprivation conditions.

Table 5. Correlation of Demographic Characteristics with Subjective Sleepiness

BASELINE CORRELATIONS				
	AGE	IQ	MEQ Score	ME (self-reported)
ME Score	0.167	-0.204		
	0.258	0.165		
ME reported	-0.235	0.095		
	0.108	0.520		
Karolinska 5	-0.176	0.091	-0.201	-0.066
	0.231	0.536	0.170	0.654
Stanford 5	-0.133	0.142	-0.117	-0.008
	0.369	0.337	0.428	0.959

SLEEP DEPRIVED CORRELATIONS				
	AGE	IQ	MEQ Score	ME (self-reported_
ME Score	0.167	-0.204		
	0.258	0.165		
ME reported	-0.235	0.095		
	0.108	0.520		
Karolinska 14	-0.069	0.221	**-0.302***	-0.106
	0.643	0.131	**0.037***	0.475
Stanford 14	-0.163	0.222	**-0.302***	-0.113
	0.268	0.130	**0.037***	0.446

Pearson correlation

P-value

Under sleep deprivation conditions, a significant negative correlation was found between MEQ score and both Karolinska (r= -.302, p=.037) and Stanford (r= -.302, p=.037) sleepiness ratings. The latter indicated that the lower a participant's MEQ score (i.e., "eveningness"), the more sleepy the participant rated him/herself during sleep deprivation. Note that the sleep deprived testing session was conducted from 0530 until 1100 in the morning and the evening types reported greater subjective sleepiness.

T-tests (α=.05) were conducted to determine if there was a significant difference in subjective sleepiness based on gender. There was no significant difference in subjective sleepiness ratings at baseline; however, the results approached significance for gender after 22 hours of sleep deprivation. Females reported higher subjective sleepiness on both Stanford (p=.054) and Karolinska scales (p=.097). These results are consistent with research indicating women generally need more sleep than men (Ferrara and DeGennaro, 2001) and are less tolerant of sleep deprivation with respect to mood (as cited in (J. Caldwell and LeDuc, 1998).

An ANOVA (α=.05) was also conducted to determine if there was a significant difference in subjective sleepiness based on occupation. Results indicated that students reported significantly higher sleepiness than military participants on the Stanford Sleepiness scale (p=.022) after 22 hours of sleep deprivation. Because there was no significant difference on the Karolinska scale (p=.185), these results were not conclusive.

4.3 Self-reported Sleepiness and Attention (PVT Performance)

To determine the relationship between self-reported sleepiness and mean reaction time (RT) on the PVT, a regression analysis was conducted. Figure 16 shows the relationship between self-reported sleepiness as rated on the Karolinska Scale and mean RT. A quadratic function best described the data and accounted for 86.9% of the variance (p=.000). The residuals were normally distributed (p>.15). Similar results were found for self-reported sleepiness as rated on the Stanford Sleepiness Scale; these results are illustrated in Figure 17. Again, a quadratic function best described the data but accounted for slightly less variance than the Karolinska Scale (75.5%) (p=.000). Residuals for the Stanford regression were also normally distributed (p=.147) (see Appendix C for analysis results). For both the Karolinska and the Stanford Sleepiness Scales, RT increased more rapidly with increasing sleepiness ratings. In short, self-reported sleepiness significantly predicted mean RT on the PVT. The Karolinska Scale accounted for more variance in the prediction model than the Stanford Sleepiness Scale, probably because the former is a 9-point scale whereas the latter is a 7-point scale. As a function of its greater range, the Karolinska Scale may be a better tool for predicting performance than the Stanford Sleepiness Scale.

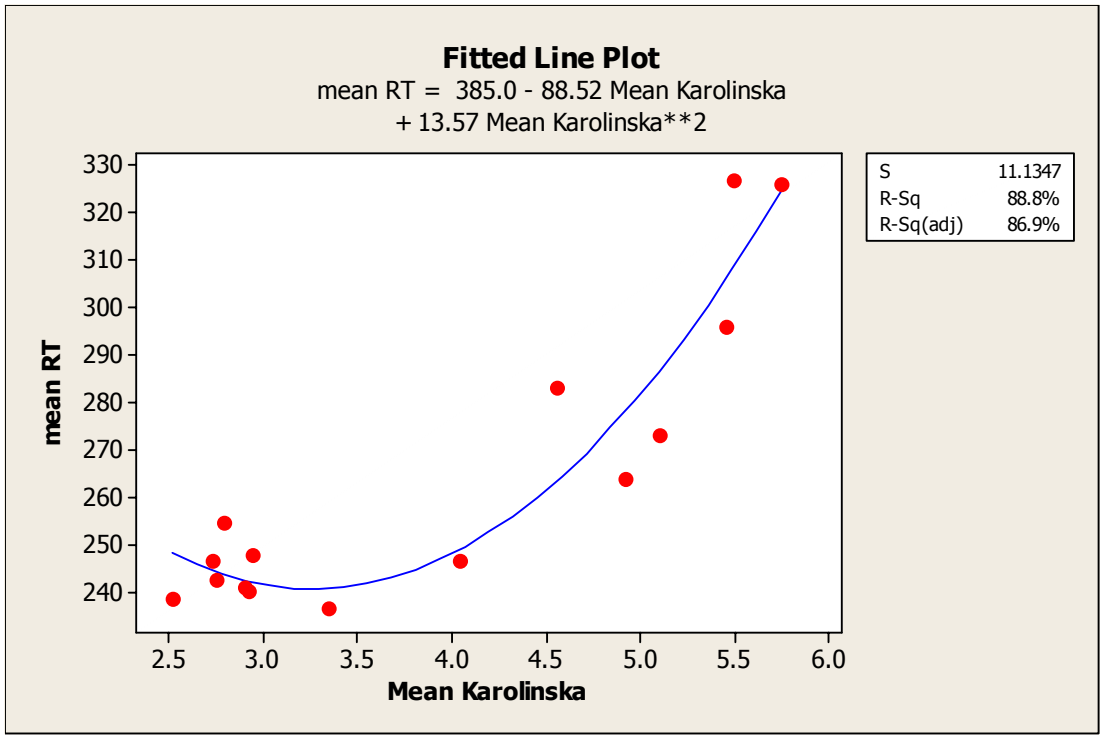


Figure 16. Mean RT vs. Mean Karolinska

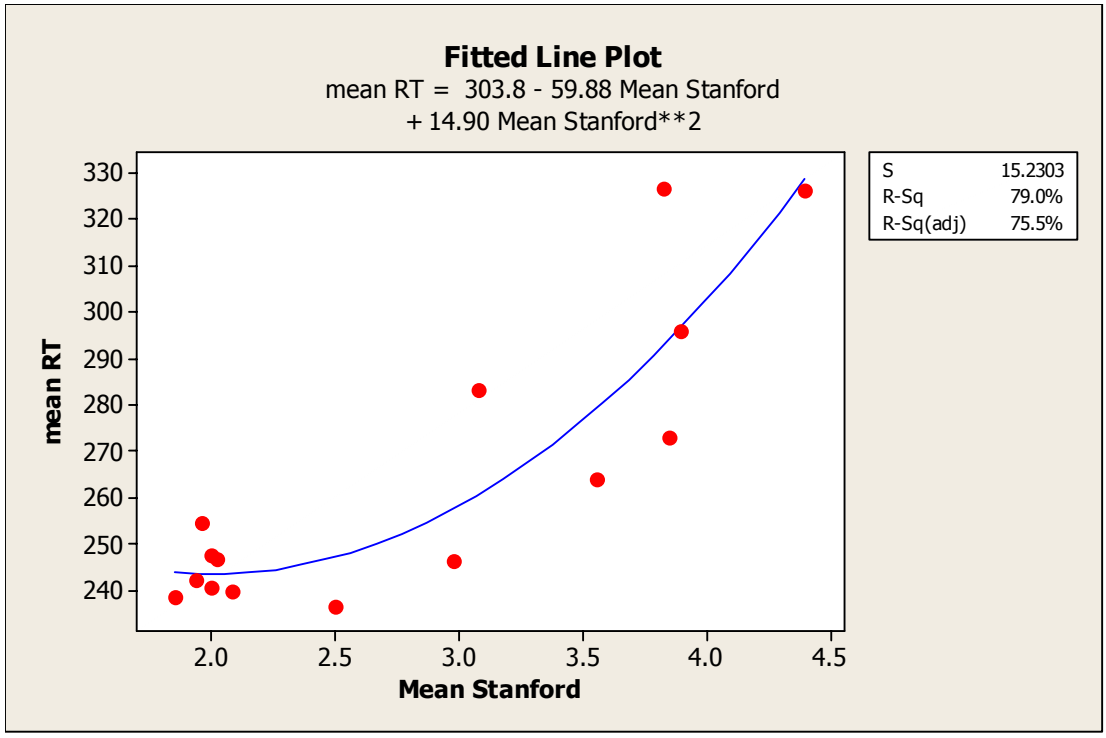


Figure 17. Mean RT vs. Mean Stanford

4.4 Cognitive Task Data Characteristics

4.4.1 Normality and Equality of Variance

Normality for all cognitive tests was tested using the Kolmogorov-Smirnov goodness of fit test ($\alpha = .05$). The Kolmogorov-Smirnov test was used because it is less sensitive to deviations from normality at the tails of the distribution (Technology, 2005). Equality of variance between baseline and sleep-deprived sessions was determined by F-test ($\alpha = .05$) for normal data and by Levene's test ($\alpha = .05$) for non-normal data. Appendix D contains all normality and equal variance test results.

4.4.2 Outliers

This dissertation analyzed data sets twice—with and without outliers included. A potential outlier was defined as a data point that was larger than two standard deviations from the mean (Montgomery, 1997). If outliers (as defined above) were present in the initial data set, the data was reanalyzed for normality and equal variance with no outliers. Because outliers in sleep deprivation data may provide valuable information about individual differences in response to sleep deprivation, data were analyzed both with and without outliers. The normality of the data was not affected by outliers for any cognitive test and the equality of variance between sessions was only affected by outliers on the Letter Sets test. Furthermore, the assumptions of the paired t-tests (i.e., normality of paired differences) were met regardless of whether outliers were included in the data set (See Table 10). Therefore, although data both with and without outliers is presented for comparison purposes, the assumptions of the analyses were met with outliers and therefore, the results were interpreted using outliers to prevent losing valuable data.

4.4.3 Data Characteristics

All cognitive tests were administered to 48 participants; however, due to computer malfunctions during data collection, the Iowa Gambling Task had 37 initial data points and the Psychomotor Vigilance Task had 47 initial data points. Table 6 summarizes the distribution and the equality of variance for baseline and sleep deprived sessions, with and without outliers (See Appendix D for normal probability plots and equal variance plots of all cognitive tests).

Table 6. Summary of Data Characteristics

	Outliers in Data			No Outliers in Data		
Cognitive Test	**Baseline Distribution**	**SD Distribution**	**Equality of Variance**	**Baseline Distribution**	**SD Distribution**	**Equality of variance**
WOMBAT	normal	normal	equal	n/a	n/a	n/a
LNS	normal	normal	equal	normal	normal	equal
LS	normal	normal	equal	normal	normal	not equal
NS	normal	normal	equal	normal	normal	equal
MTT	normal	normal	equal	normal	normal	equal
PVT	normal	not normal	not equal	normal	not normal	not equal
IGT	normal	normal	equal	normal	normal	equal

With the exception of the PVT, all data sets were normally distributed. The variance between sessions was not equal for the PVT or the Letter Sets Test without outliers in the data.

4.5 Effects of Sleep Deprivation on Attention

The effects of sleep deprivation on attention were analyzed using the Psychomotor Vigilance Task (PVT). The PVT was administered every two hours for a total of 15 sessions. The baseline PVT and sleep-deprived PVT sessions (PVT session 5 and 14, respectively) were chosen to represent baseline and sleep-deprived data because they were administered during the same block of time as the other cognitive tests at 9 hours SD and 27 hours SD, respectively (See Appendix A for testing schedule). Mean reaction time (mean RT) was significantly degraded by sleep deprivation ($p=.000$). Figure 18 shows the mean reaction time for both baseline and sleep deprived sessions—260.5 ms and 346.9 ms, respectively. Session 5 (i.e., Baseline) PVT data were normally distributed ($p=.06$) and Session 14 (i.e., sleep deprived) PVT data were not normally distributed ($p<.01$). The sessions did not have equal variance ($p=.000$) (see Appendix D).

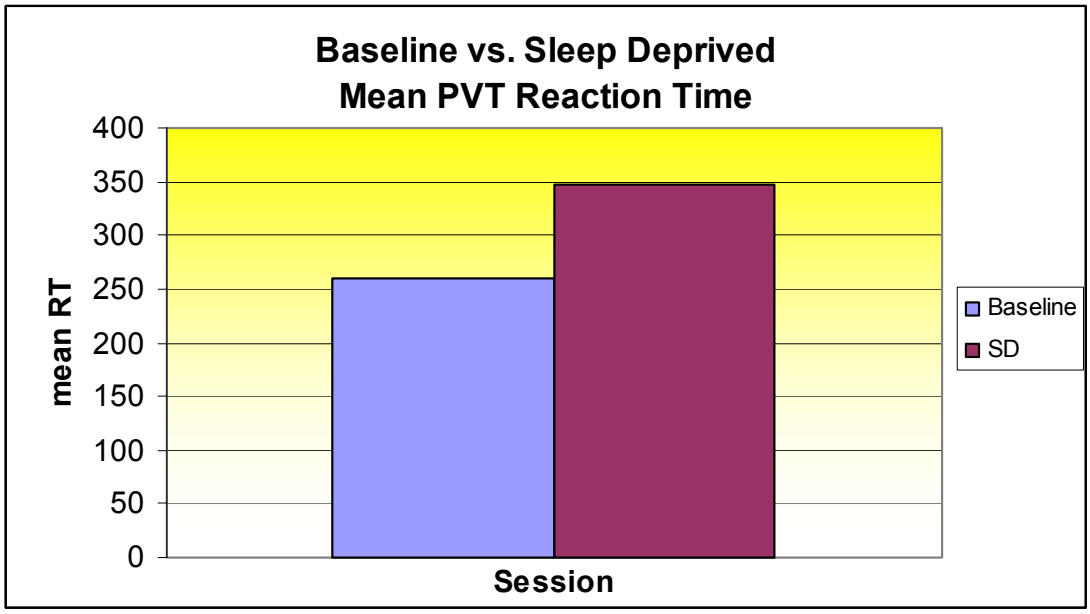


Figure 18. Mean PVT Reaction Time by Session

There were four outliers—two in the baseline data and two in the sleep deprived data. The baseline session data values (413.01, 364.24) were further than two standard deviations from the baseline mean (M = 260.5, SD = 46.6) and the two sleep-deprived data values (703.8, 888.4) were further than two standard deviations from the sleep-deprived mean (M = 346.9, SD = 135.5). The outliers were removed from both sessions and the data subset was then reanalyzed for normality and equal variance (N=46, N=46). There was no change in the data characteristics. The baseline data were normally distributed (p>.15); sleep-deprived data were not normally distributed (p=.048); and the variance was not equal between sessions (p=.000) (See Appendix D).

Figure 19 shows the differences between baseline and sleep deprived scores for each participant. Six participants improved with sleep deprivation and 40 degraded with sleep deprivation. The mean improvement in score was 10.2 (SD=9.1). The mean degradation in score was 98.0 (SD=120.0). The mean baseline score for those participants who improved with sleep deprivation was 248.5 (SD=14.8) while the mean baseline score for those participants who degraded was 262.3 (SD=49.6). There was no significant difference in baseline scores for those participants who improved compared to those who degraded.

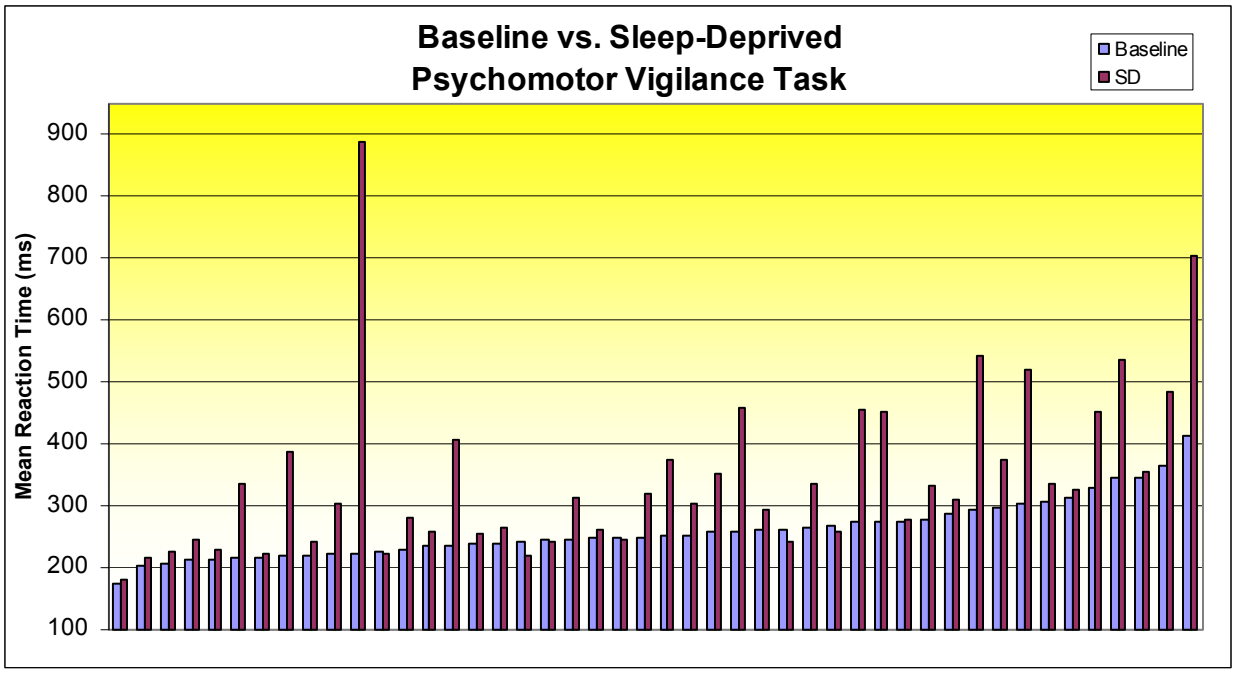


Figure 19. Psychomotor Vigilance Task Performance by Session

Further analysis of PVT indicated there were significant effects of SD over time. Figure 20 shows the mean reaction time (RT) over fifteen intervals up to 30 hours of total sleep loss.

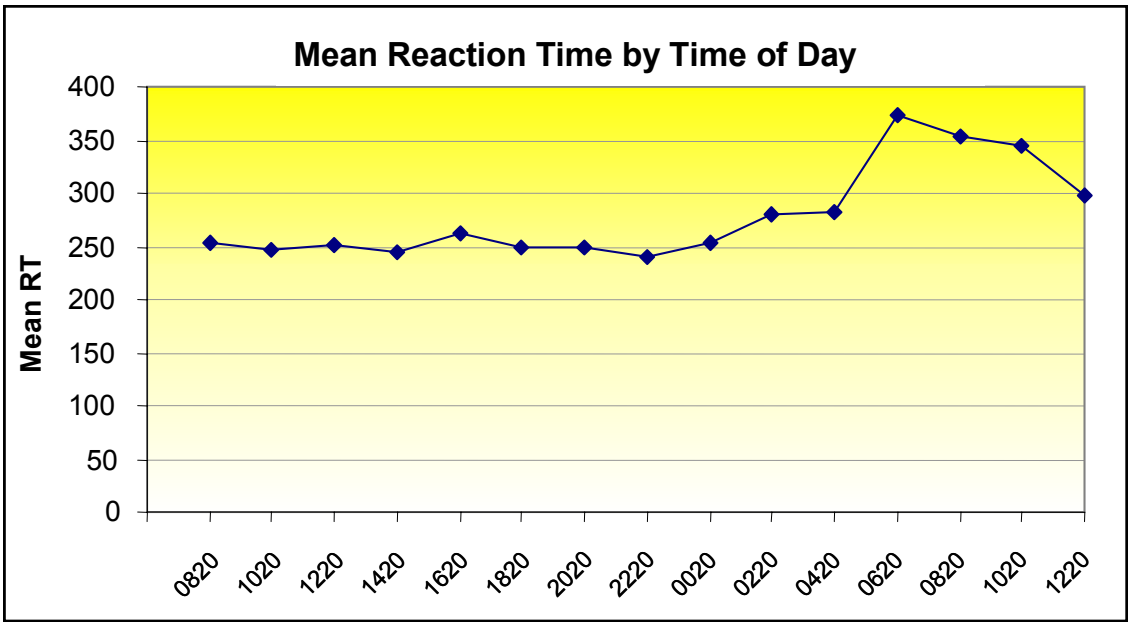


Figure 20. Mean RT by Session

A regression analysis of PVT data indicated a significant linear, quadratic, and cubic effects of SD on mean RT (See Appendix H for results of linear and quadratic regressions). The best fit (as determined by adjusted R^2 value) was the cubic regression equation: Mean RT = 271.1 - 19.24 session + 2.937 session^2 - 0.09965 session^3. Figure 21 shows the cubic fit of the data. The equation accounted for 65.1% of the variance in the data (p=.002). Table 7 shows the ANOVA results for the cubic regression. It is likely that the combined effects of sleep deprivation and time of day resulted in this cubic function.

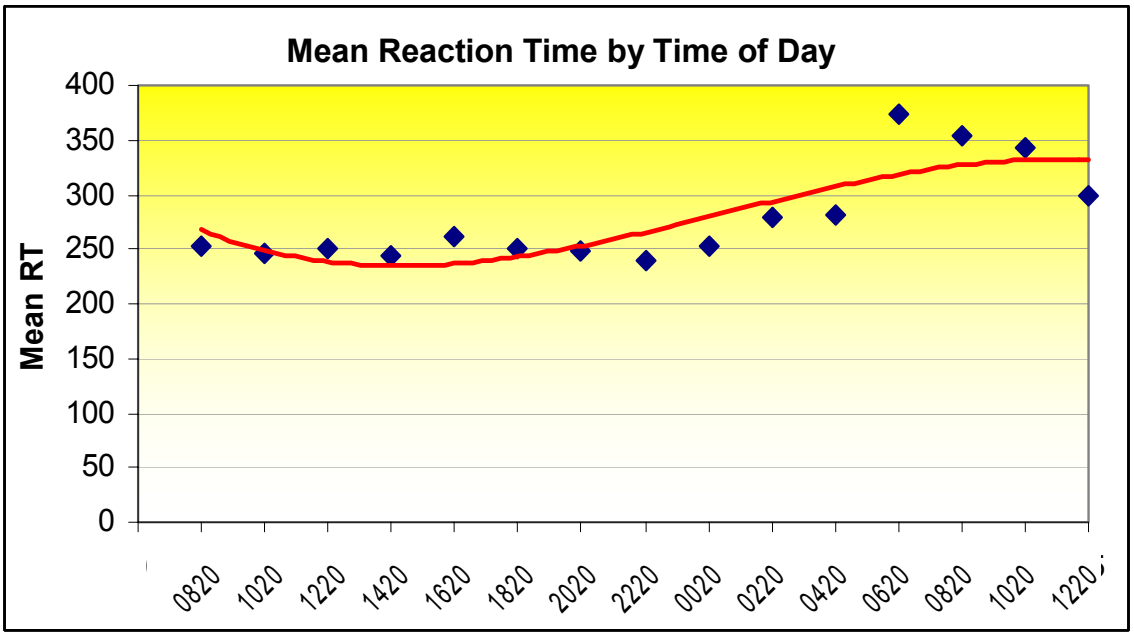


Figure 21. Cubic regression of PVT data

Table 7. ANOVA for Cubic Regression of Mean RT

Source	DF	SS	MS	F	P
Regression	3	9612.3	3204.11	9.69	0.002
Error	11	3636.7	330.61		
Total	14	13249.0			

A one-way ANOVA for session was conducted to determine which PVT sessions had significantly different mean RT scores. Outliers were included in the analysis since it was assumed that outliers represented sleep deprivation effects. The sphericity assumption was not

met; therefore, a Geisser-Greenhouse correction was used. Table 8 shows the results of the ANOVA with outliers in the data.

Table 8. ANOVA of Mean RT over Interval

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
session	Sphericity Assumed	1245386.08	14	88956.149	6.614	.000
	Greenhouse-Geisser	1245386.08	1.970	632245.046	6.614	.002
Error(session)	Sphericity Assumed	8473723.43	630	13450.355		
	Greenhouse-Geisser	8473723.43	88.640	95596.766		

Results indicated a significant change in mean RT across sessions (i.e., time of day and hours of sleep deprivation) ($p=.003$). Post hoc paired comparisons were made using a Tukey's HSD correction ($\alpha=.05$) (See Appendix H); however, no significant comparisons were found (likely a function of the fact that the Tukey is a conservative test). Means were then compared using Bonferroni-corrected paired comparisons ($\alpha=.05$) to determine if a more robust test yielded different results (See Appendix H). Table 9 shows the Bonferroni results. Significant differences between intervals are denoted by an X.

Using the less conservative Bonferroni correction, significant differences in mean RT between sessions were revealed. Because the ANOVA was significant for session and because of the exploratory nature of this analysis, it was concluded that a less conservative approach was warranted and the Bonferroni results were used for discussion purposes.

Cochran-Mantel-Haenszel (CMH) chi square tests were conducted on various demographic variables to determine whether any significant differences were revealed between those participants whose mean RT improved after sleep deprivation and those whose mean RT degraded after sleep deprivation. There were no demographic factors that significantly differentiated the two groups (See Appendix D).

Table 9. Significant Mean RT Differences by Session

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1														X	
2											X			X	
3														X	
4					X						X			X	X
5				X				X						X	
6											X			X	
7														X	
8					X						X			X	X
9														X	
10															
11		X		X		X		X							
12															
13															
14	X	X	X	X	X	X	X	X	X						
15				X				X							

By virtue of the number of published studies in which the PVT has been used to quantify sleep loss effects, the PVT is becoming the gold standard for measuring the effects of SD on attention. Results of this study supported the hypothesis that attention would be degraded by sleep deprivation. The results of the paired t-test indicated a significant increase in mean RT due to sleep deprivation, and the significant session effect found across sessions using ANOVA suggested a time of day effect as well. Thus, these results are consistent with previous studies in which degraded attention (as measured by mean RT) in response to sleep deprivation has been shown (Himashree et al., 2002; Giam, 1997; Wesensten, 2003). In this study, those demographic factors which were measured did not account for individual differences in resilience to sleep deprivation (based on PVT performance). Results of the PVT analysis established that participants were indeed sleep deprived enough to cause a degradation in lower-order cognitive capabilities; the latter findings served as the foundation for conducting further analyses on the effects of sleep deprivation on working memory and executive function.

4.6 Effects of Sleep Deprivation on Tests of Working Memory and Executive Function

It was hypothesized that tasks requiring rule-based decision making would not be degraded by sleep deprivation but that tasks requiring working memory, planning, decision making, and situation awareness would be degraded by sleep deprivation. To test the above

hypothesis, paired t-tests were used to determine if there was a significant difference between baseline and sleep-deprived performance on several cognitive tasks. Assumptions for the paired t-test include the following: 1) independence of paired differences, and 2) normality of paired differences (Hollander and Wolfe, 1999). To determine whether the data might violate assumptions for each paired t-test, the normality of paired differences was tested. It was assumed that all samples were independent. To prevent the loss of potentially important data, analyses were performed both with and without outliers in the data sets. Table 10 shows the normality results for paired differences on each cognitive test (See Appendix D for normal probability plots).

Table 10. Normality of Paired Differences for Cognitive Test Results

Cognitive Test	Normality of Paired Differences (p-value)	Normality of Paired Differences with no outliers (p-value)
WOMBAT	Normal (p=.069)	Normal (p=.069)
LNS	Normal (p>.15)	Normal (p>.15)
LS	Normal (p>.15)	Normal (p>.15)
NS	Normal (p>.15)	Normal (p>.15)
MTT	Normal (p>.15)	Normal (p>.15)
IGT	Not normal (p<.01)	Not normal (p<.01)
PVT	Not normal (p<.01)	Not normal (p<.01)

The above results suggested that the paired t-test was not appropriate for either the PVT or IGT because both data sets violated the normality assumption ($p<.01$) and the distributions were not symmetrical (PVT skewness = -0.93, IGT skewness = 0.87). Therefore, the Wilcoxon paired sign rank test was used in addition to the parametric test to analyze the PVT and IGT data because the Wilcoxon paired signed rank test does not assume normality or a symmetrical distribution (Hollander and Wolfe, 1999). Additionally, the initial PVT data set was normalized using a log transform to allow a paired t-test analysis; however, because the log PVT paired differences were not normally distributed ($p<.01$) and thus still violated the assumptions of the paired t-test, the log PVT analysis was not used (See Appendix D).

4.6.1 Paired T-Test Results

Table 11 shows the results of the paired t-tests for each cognitive test.

Table 11. Significance of Paired T-tests

Cognitive Test	Outliers in Data	No Outliers in Data
WOMBAT	(t = -1.15, p = 0.257)	(t = -1.15, p = 0.257)
LNS	(t = 0.15, p = 0.879)	(t = -0.41, p = 0.681)
LS	(t = -1.18, p = 0.245)	(t = -0.89, p = 0.380)
NS	(t = 1.09, p = 0.232)	(t = 0.96, p = 0.342)
MTT	(t = -4.59, p = 0.000)*	(t = -4.85, p = 0.000)*
IGT parametric	(t = -1.26, p = 0.214)	(t = -1.56, p = 0.127)
IGT nonparametric	(t = -1.886, p = 0.059)	(t = -1.591, p = 0.112)
PVT parametric	(t = -4.84, p = 0.000)*	(t = -5.53, p = 0.000)*
PVT nonparametric	(t = -4.886, p = 0.000)*	(t = -4.27, p = 0.000)*

*significant at $\alpha = .05$

With outliers in the data sets, results indicated a significant difference between baseline and sleep-deprived performance on the MTT and PVT ($p < .05$). The IGT nonparametric results approached significance ($p=0.059$). When outliers were excluded from the data, there was still a significant difference on the MTT and PVT ($p < .05$). The nonparametric test for IGT approached significance with outliers in the data ($p=.059$), but the parametric test was not significant ($p=.214$). The parametric and nonparametric results for both PVT and IGT were consistent (i.e., PVT was significant; IGT was not significant) when outliers were excluded from the data. The following sections describe the results for each cognitive test in detail.

4.6.2 Working Memory (Letter Number Sequencing)

Paired t-test results indicated that working memory was not degraded by sleep deprivation (as measured by the number of correct answers on the Letter Number Sequencing (LNS) task ($p=.879$). Figure 22 shows the mean number of correct LNS responses for both baseline and sleep deprived sessions. The mean baseline score was 8.5 (SD=1.9) and the mean SD score was 8.45 (SD=1.8).

Data for LNS under both baseline and sleep deprivation conditions were normally distributed ($p>.15$) and had equal variance ($p=.648$) (see Appendix D for normal probability plots and equal variance plots). Two outliers were found in the baseline data. Both data values

(13, 15) were further than two standard deviations from the mean (M = 8.5, SD = 1.9); therefore, both outliers were removed. The data subset was then reanalyzed for normality and equal variance (N=46). Reanalysis of the LNS data without outliers indicated normally distributed data for both baseline (p>.15) and sleep deprived sessions (p>.15) as well as equal variance for both sessions (p=.427).

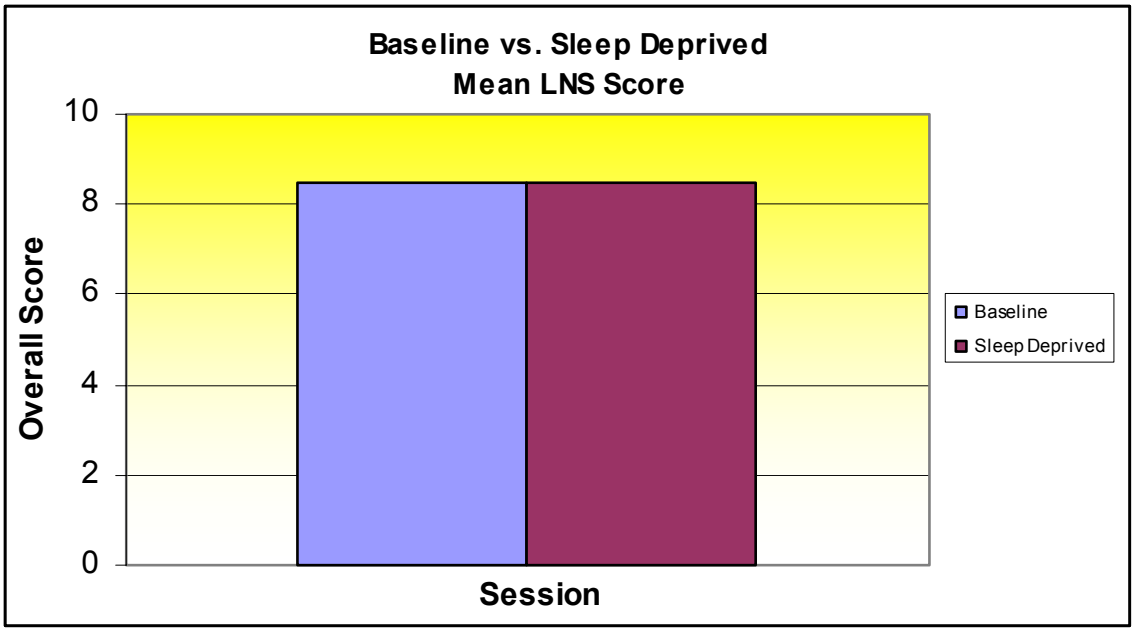


Figure 22. Mean LNS Score by Session

Figure 23 shows number of correct items on LNS for baseline and sleep deprived conditions for each participant. Participants are rank-ordered by baseline score. Eighteen participants improved with sleep deprivation, nine participants showed no difference in score, and 11 degraded with sleep deprivation. The mean baseline score for those participants who improved with sleep deprivation was 7.6 (SD=1.2) while the mean baseline score for those participants who degraded was 9.9 (SD=1.9). The mean baseline score of those whose performance did not change with sleep deprivation was 7.1 (SD=0.6).

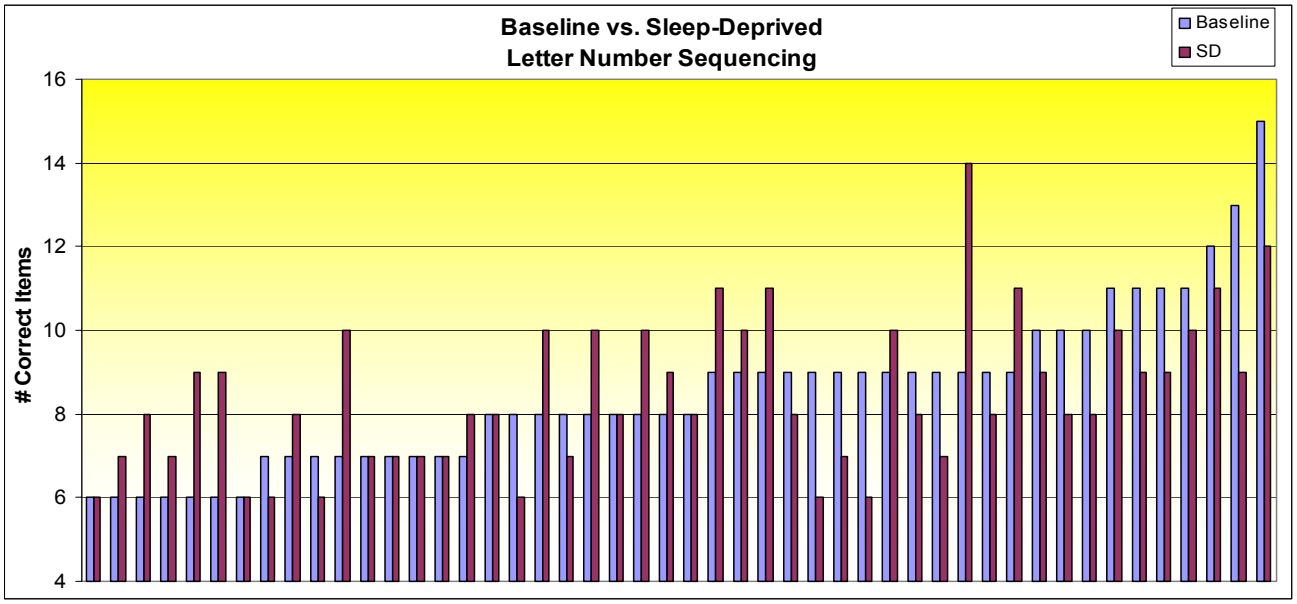


Figure 23. LNS Scores by Session

Participants whose performance degraded with sleep deprivation had significantly higher baseline scores than those whose performance improved (p=.000) and those whose performance was unchanged (p=.000) (the latter two groups did not differ, p=.140). The mean improvement in score was 1.8 items (SD=0.9) while the mean degradation was 1.9 items (SD=1.05). Cochran-Mantel-Haenszel (CMH) statistics indicated that there were no significant general association of demographic characteristics with participants who improved compared to those who degraded (See Appendix D).

Results of this dissertation did not support the hypothesis that working memory (as measured by LNS performance) would be degraded by sleep deprivation. Although the literature search failed to find previous studies in which LNS was used in a sleep deprivation study, the LNS has been used to study working memory dysfunction associated with schizophrenia (Boulay, 2004; Kent, Fox, Michie, and Jablensky, 2004; McGurk et al., 2004), ADHD (Hanford, 2001), and traumatic brain injury (Donders, Tulsky, and Zhu, 2001), and it is a subtest of the WAIS-III. LNS includes elements of working memory processes (as cited in Haut et al., 2000) and purportedly measures more than simple digit span (Crow, 2000). Despite some evidence for degraded working memory in some sleep deprivation studies (Thomas, Sing, et al, 2000), results from this study (using LNS) are consistent with other studies in which working memory was

sustained during sleep deprivation (Linde and Bergstrom, 1992; Wimmer, Hoffman, et al., 1992; Glenville, Broughton, Wing, and Wilkinson, 1978).

4.6.3 Inductive Reasoning (Letter Sets)

Inductive reasoning (as measured by number of correct responses on the Letter Sets test) was not degraded by sleep deprivation (p=.245). Figure 24 shows the mean number of correct LS answers for both baseline and sleep deprived sessions. Data for both baseline and sleep deprivation conditions (N=48) were normally distributed (p>.15) and had equal variance (p=.400) (see Appendix D).

There was one outlier in the baseline data. The data value (3) was further than two standard deviations from the mean (M = 10.3, SD = 2.7); therefore, the outlier was removed the data. The data subset was then reanalyzed for normality and equal variance (N =47). Reanalysis without the outlier indicated that data from both sessions were normally distributed (p>.15) and had equal variance (p=.870) (See Appendix D).

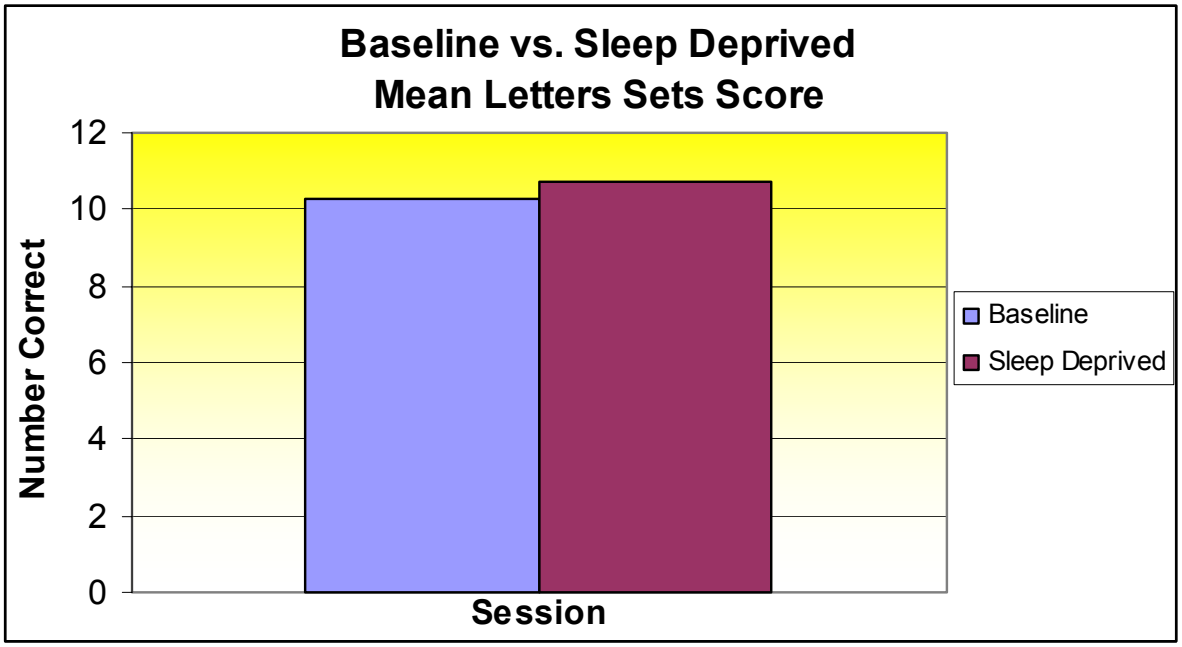


Figure 24. Mean Letter Sets Score by Session

Figure 25 shows participant scores for baseline and sleep deprivation conditions rank-ordered by baseline score. Participants are rank-ordered by baseline score. Twenty-two participants improved with sleep deprivation, seven participants did not change, and 19 degraded

with sleep deprivation. The mean improvement in score was 2.8 (SD=1.4). The mean degradation in score was 2.1 (SD=1.2).

The mean baseline score for those participants who improved with sleep deprivation was 8.8 (SD=2.6) while the mean baseline score for those participants who degraded was 11.9 (SD=2.3). The mean baseline score of those who showed no difference was 10.4 (SD=1.7).

Participants whose performance degraded with sleep deprivation had significantly higher baseline scores than those whose performance improved (p=.000). There was no significant difference in baseline scores between those whose score did not change versus those who improved (p=.067), and between those whose score did not change versus those who degraded (p=.062). The mean improvement in score was 2.8 points (SD=1.4) while the mean degradation was 2.1 points (SD=1.2). Analyses of demographic variables failed to reveal any significant differences between participants who improved and those who degraded with sleep deprivation (See Appendix D).

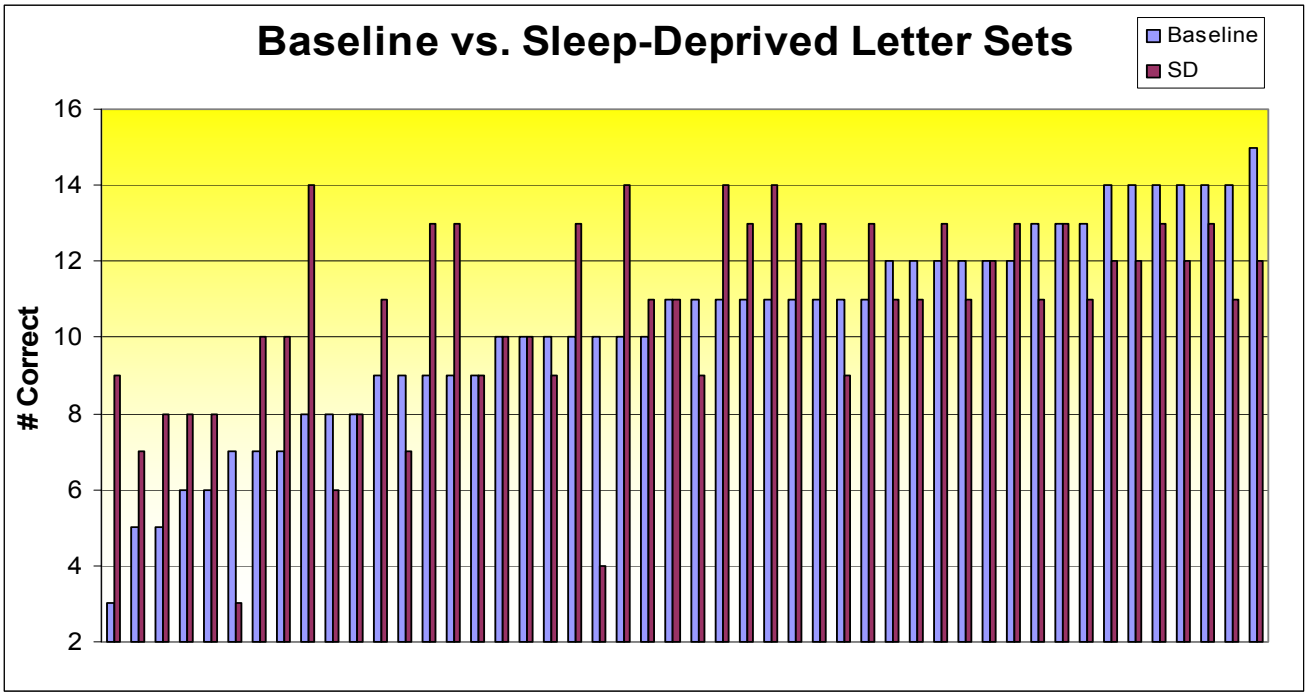


Figure 25. Letter Sets Scores by Session

As hypothesized, results of the present study showed no degradation in inductive reasoning (as measured by Letter Sets performance). However, results from previous studies are mixed, suggesting that inductive reasoning is not universally degraded by sleep deprivation or

that sleep deprivation effects on so-called "inductive reasoning" are task-specific. The results of this study were consistent with May and Kline's (1987) study in which no degradation in either logical or inductive reasoning was found using the same Nonsense Syllogisms and Letter Sets tests from the Kit of Factor-referenced Cognitive Tests (Ekstrom et al., 1976).

4.6.4 Deductive Reasoning (Logical Reasoning)

Paired t-test results failed to reveal any degradation in deductive reasoning (as measured by number of correct responses on the Nonsense Syllogisms test) (p=.232). Figure 26 shows the number of correct NS answers for both baseline and sleep deprived sessions. Both baseline and sleep deprived data were normally distributed (p>.15) and had equal variance (p=.126) (See Appendix D).

There were two outliers in the data (one in each session). The baseline data value (12) was further than two standard deviations from the baseline mean (M = 4.7, SD = 2.7) and the sleep deprived data value (14) was further than two standard deviations from the baseline mean (M = 7.4, SD = 3.2); therefore, the outliers were removed the data. The data subset was then reanalyzed for normality and equal variance (N=47). Reanalysis of the NS data indicated that data from both sessions were normally distributed (p>.15). Without outliers in the data, the two sessions have equal variance (p=.077) (See Appendix D).

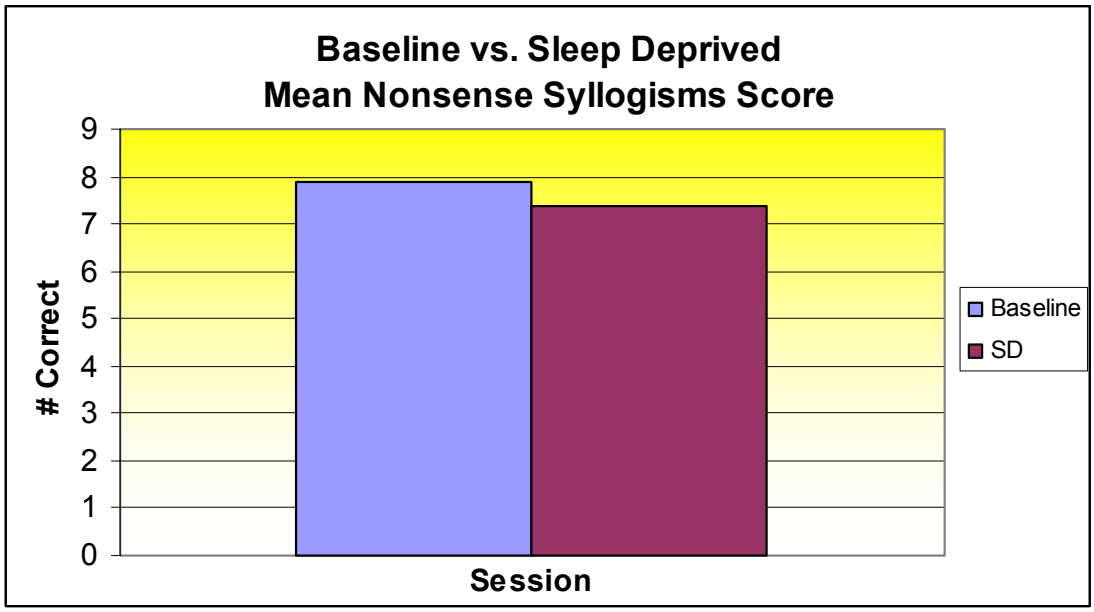


Figure 26. Mean NS Score by Session

Figure 27 shows participant scores for baseline and sleep deprivation conditions, rank ordered by baseline performance. Nineteen participants improved with sleep deprivation, five participants did not change with sleep deprivation, and 24 participants degraded with sleep deprivation. The mean improvement in score was 2.5 (SD=1.2). The mean degradation in score was 3.0 (SD=2.0). The mean baseline score for those participants who improved with sleep deprivation was 7.5 (SD=2.6) while the mean baseline score for those participants who degraded was 8.3 (SD=2.5). The mean baseline score of those whose performance did not change was 7.2 (SD=2.3).

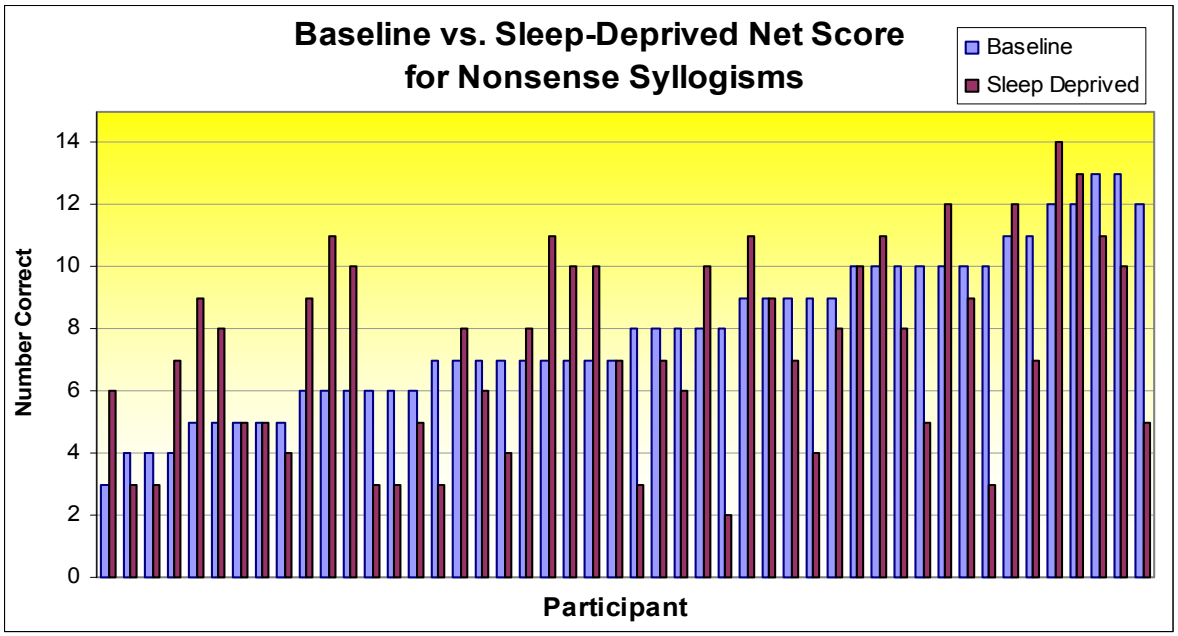


Figure 27. Logical Reasoning Scores by Session

No significant differences in baseline scores were found among participants who improved, degraded, or stayed the same. Participants who improved after sleep deprivation had significantly higher WASI IQ scores than participants who degraded (p=.031). However, this difference was no longer significant when WASI IQ was covaried for education and language. The CMH measure of general association indicated a significant association between improvement vs. degradation and occupation (p=.017) as well as morningness-eveningness (p=.008) (See Appendix D).

The results of the deductive reasoning test (Nonsense Syllogisms) supported the hypothesis that rule-based decision making would not be degraded by sleep deprivation. The results also are consistent with results of previous studies in which no degradation in logical reasoning was found (Harrison and Horne, 1999; Linde and Bergstrom, 1992; Ryman, Naitoh, and Englund, 1985).

4.6.5 Planning (Maze Tracing Test)

Results of the paired t-test indicated a significant improvement in planning in response to SD (as measured by the number of mazes completed in 3 minutes) ($p=.000$). Figure 28 shows the number of mazes completed in 3 minutes for baseline and sleep deprivation conditions. Normal probability plots of the MTT data (N=48) indicated that both baseline and sleep-deprived MTT data were normally distributed ($p>.15$). The sessions had equal variance ($p=.262$) (see Appendix D).

There were three outliers in the data—one outlier in the baseline data and two outliers in the sleep-deprived data. The baseline session data value (20) was further than two standard deviations from the baseline mean ($M = 12.0$, $SD = 3.1$) and the two sleep-deprived data values (23, 24) were further than two standard deviations from the sleep-deprived mean ($M = 13.4$, $SD = 3.6$). One subject had two outliers (one each from baseline and sleep-deprived sessions); therefore, two subjects' data were removed from both sessions. The data subset was then reanalyzed for normality and equal variance (N=46, N=47). Reanalysis of the data without outliers indicated that both sessions were normally distributed ($p>.15$) and that the sessions had equal variance ($p=.637$) (see Appendix D).

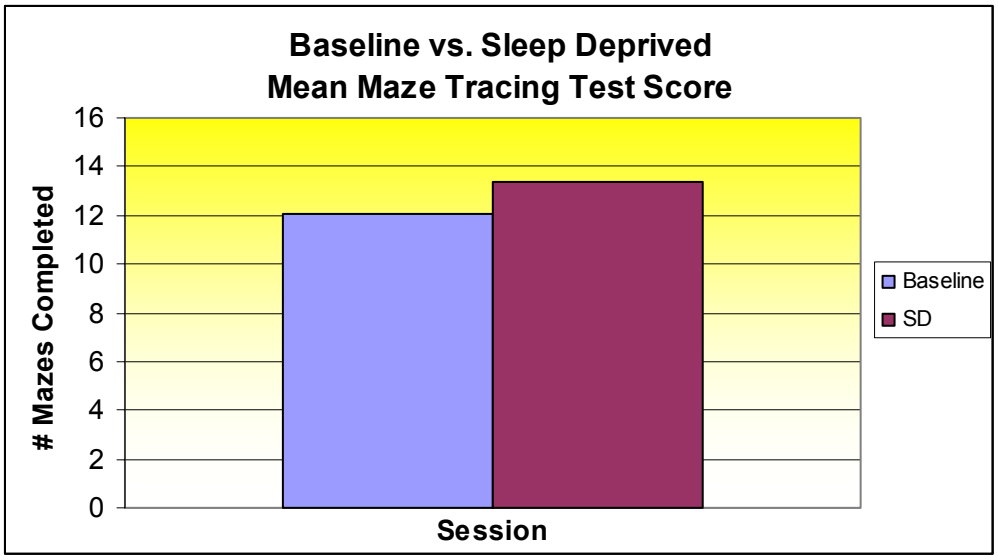


Figure 28. Mean MTT Scores by Session

Figure 29 shows the differences between baseline and sleep deprived scores for each participant. Participants were rank-ordered by baseline score. Thirty-two participants improved with sleep deprivation, six participants showed no difference in score, and 10 degraded with sleep deprivation. The mean improvement in score was 2.4 (SD=1.4). The mean degradation in score was 1.4 (SD=0.7).

The mean baseline score for those participants who improved with sleep deprivation was 12.0 (SD=3.3) while the mean baseline score for those participants who degraded was 12.8 (SD=2.5). The mean baseline score of those who showed no difference was 11.2 (SD=2.9). There was no significant difference in baseline scores for those participants who improved, degraded, or stayed the same. The CMH measure of general association indicated a significant association between improvement vs. degradation and ethnicity (p=.044) (See Appendix D).

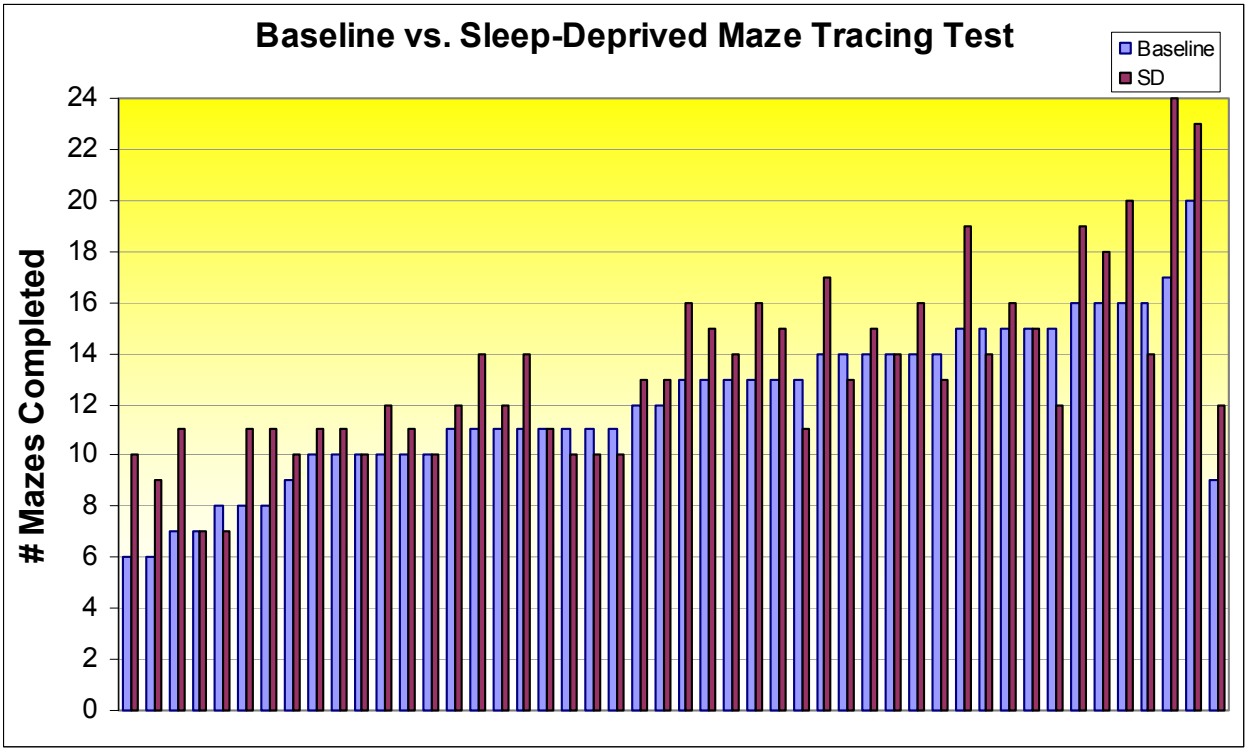


Figure 29. Maze Tracing Test Scores by Session

The results of this study did not confirm the results of previous studies in which planning was degraded by sleep deprivation (Harrison and Horne, 1999; Wimmer, Hoffman et al., 1992). In the Harrison and Horne studies, a task called "Master Planner" was used to measure planning. In the present study, the maze tracing test was used as the measure of planning. A review of the literature failed to reveal prior sleep deprivation studies in which a maze test was used as a measure of planning; however, a maze test was used in this study because these types of tests are commonly used to study spatial planning processes in both normal and brain-damaged populations (Karnath and Wallesch, 1992; Karnath, Wallesch, and Zimmermann, 1991); and maze tests are an accepted task of planning in the neuropsychology community. Contrary to expectations, results showed a significant improvement in maze tracing performance from baseline (rested) to sleep-deprived. This improvement suggests that the maze task had a substantial learning (practice) component.

4.6.6 Post Hoc Analyses for Reasoning and Planning

Because the dependent variables proposed in the method section failed to yield significant results on inductive reasoning (Letter Sets) and deductive reasoning (Nonsense Syllogisms) tests, additional dependent variables were analyzed using paired t-tests. In addition, because the results showed a significant improvement on the planning (Maze Tracing Test) task and thus, a potential learning effect, an additional dependent variable was analyzed using a paired t-test. Net score (defined as the number of correct responses minus the number of incorrect responses) was analyzed for both LS and NS. The number of errors (defined as number of dead ends reached and/or number of retraced paths and/or number of lines crossed) on the MTT was also analyzed.

Table 12 shows the normality of paired differences for the additional LS, NS, and MTT dependent variables both with and with out outliers. The paired t-test assumptions were met for each new dependent variable (See Appendix E). Table 13 summarizes the results of the paired t-tests.

Table 12. Normality of Paired Differences

Dependent Variable	Normality of Paired Differences (p-value)	Normality of Paired Differences with no outliers (p-value)
LS net score	Normal (p>.15)	Normal (p>.15)
NS net score	Normal (p>.15)	Normal (p>.15)
MTT Errors	Normal (p>.15)	Normal (p>.15)

Table 13. Significance of paired t-tests

Dependent Variable	Outliers in Data	No Outliers in Data
LS net score	(t=-1.25, p=.217)	(t=-1.86, p=.070)
NS net score	(t=.40, p=.694)	(t=.67, p=.504)
MTT Errors	(t=-1.59, p=.119)	(t=-1.50, p=.140)

When outliers were included in the data, the results of the paired t-tests indicated no significant difference in performance between baseline and sleep deprived conditions for any of the additional dependent variables. When outliers were excluded from the data, LS net score approached significance (p=.070). However, the LS net score improved after sleep deprivation rather than degraded. The following sections explain the results of the paired t-tests in detail.

4.6.6.1 LS Net Score

Results did not indicate a significant degradation in inductive reasoning when outliers were included in the data; however, when outliers were excluded from the data, the results of the paired t-test approached significance (p=.070) indicating a trend toward significantly improved rather than degraded performance. Figure 30 shows the LS net score for both baseline and sleep deprived sessions. Participants are rank-ordered by baseline score.

The baseline data were normally distributed (p>.15), but the sleep deprived data were not normally distributed (p=.036). Baseline and sleep-deprived sessions had equal variance (p=.440) (see Appendix D). There were four outliers in the data—two in the baseline data and two in the sleep deprived data. The data values (-2, 0) were further than two standard deviations from the mean (M = 8.6, SD = 4.0) in the baseline data and the data values (-9, -2) were further than two standard deviations from the mean (M = 9.3, SD = 4.1) in the sleep deprived data; therefore, the outliers were removed the data. The data subset was then reanalyzed for normality and equal variance (N=46). Reanalysis of the LS net score data indicated that data from both sessions were normally distributed (p>.15) and the two sessions had equal variance (p=.058) (See Appendix D).

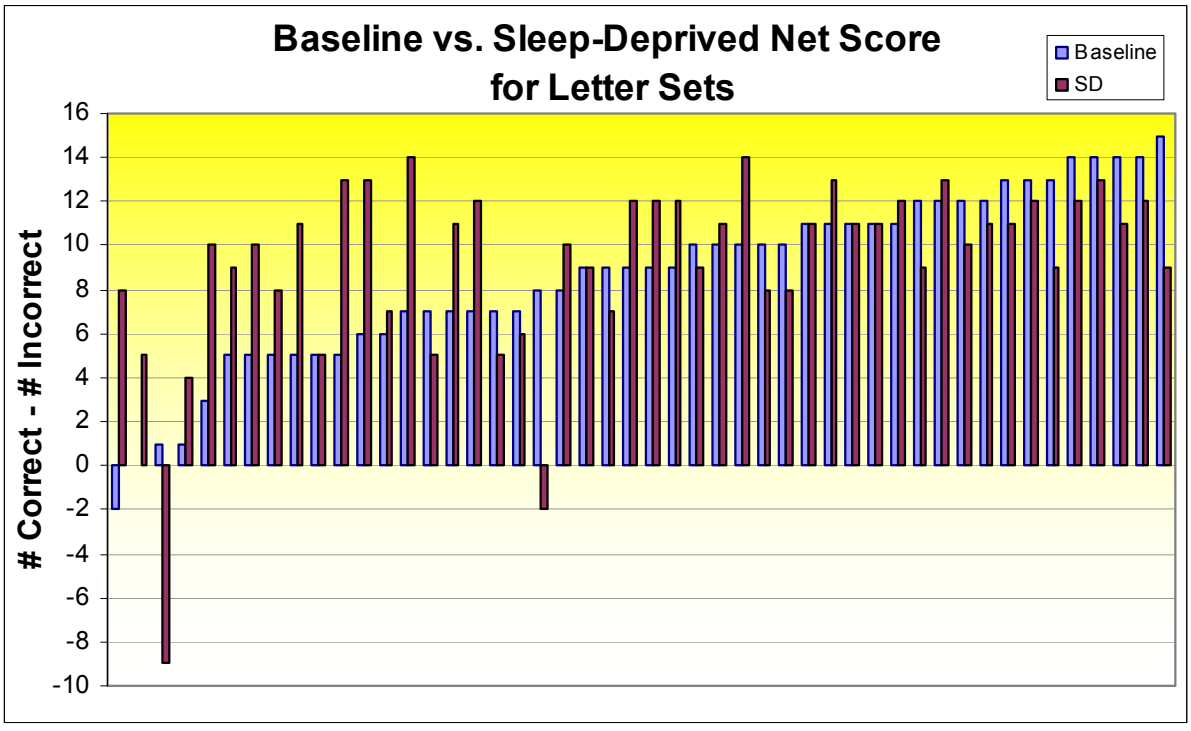


Figure 30. Net Score for Letters Sets

The results of the post hoc test for LS net score were consistent with the paired t-test results using total number of correct responses and did not indicate a degradation of inductive reasoning.

4.6.6.2 NS Net Score

Results of the post hoc analysis for NS net score did not show a significant degradation in deductive reasoning (p=.694). Figure 31 shows the NS net score for both baseline and sleep deprived sessions. Participants are rank-ordered by baseline score.

Both the baseline and sleep deprived data were normally distributed (p>.15). Baseline and sleep-deprived sessions had equal variance (p=.156) (see Appendix D). There were two outliers in the data—one in the baseline data and one in the sleep deprived data. The data values (-7) were further than two standard deviations from the mean (M = 2.7, SD = 4.15) in the baseline data and the data values (-13) was further than two standard deviations from the mean (M = 2.4, SD = 5.1) in the sleep deprived data; therefore, the outliers were removed the data. The data subset was then reanalyzed for normality and equal variance (N=47). Reanalysis of the NS net score data indicated that data from both sessions were normally distributed (p>.15). With not outliers in the data, the two sessions have equal variance (p=.134) (See Appendix D).

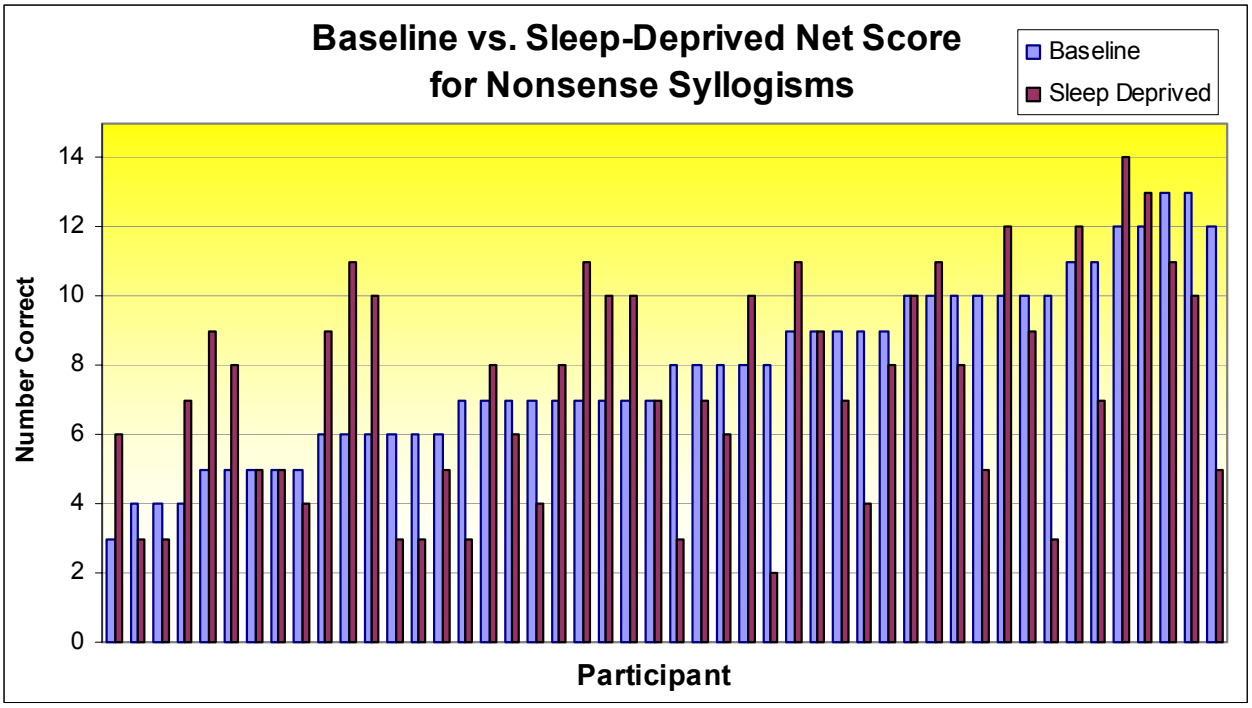


Figure 31. Net score for Nonsense Syllogisms

The post hoc analysis results for NS net score were consistent with the previous analysis of deductive reasoning which found no degradation of deductive reasoning as a result of SD.

4.6.6.3 MTT Errors

Results of the paired t-test for MTT errors did not indicate a significant increase in errors after SD (p=.119). Figure 32 shows the number of MTT errors for both baseline and sleep deprived sessions. Participants are rank-ordered by baseline score.

Baseline data were normally distributed with p=.128 and sleep deprived data were normally distributed with p=.099 (N=48). Baseline and sleep-deprived sessions had equal variance (p=.149) (see Appendix D). There were two outliers in the baseline data and three outliers in the sleep deprived data. The data values (12, 13) were further than two standard deviations from the mean (M = 4.2, SD = 3.3) in the baseline data and the sleep deprived data values (14, 17, 18) were further than two standard deviations from the mean (M = 5.7, SD = 4.1) in the sleep deprived data; therefore, the outliers were removed the data. The data subset was then reanalyzed for normality and equal variance (N=46, N=45). Reanalysis of the LS data indicated that data from both sessions were normally distributed (p>.15) and the two sessions have equal variance (p=.690 (See Appendix D).

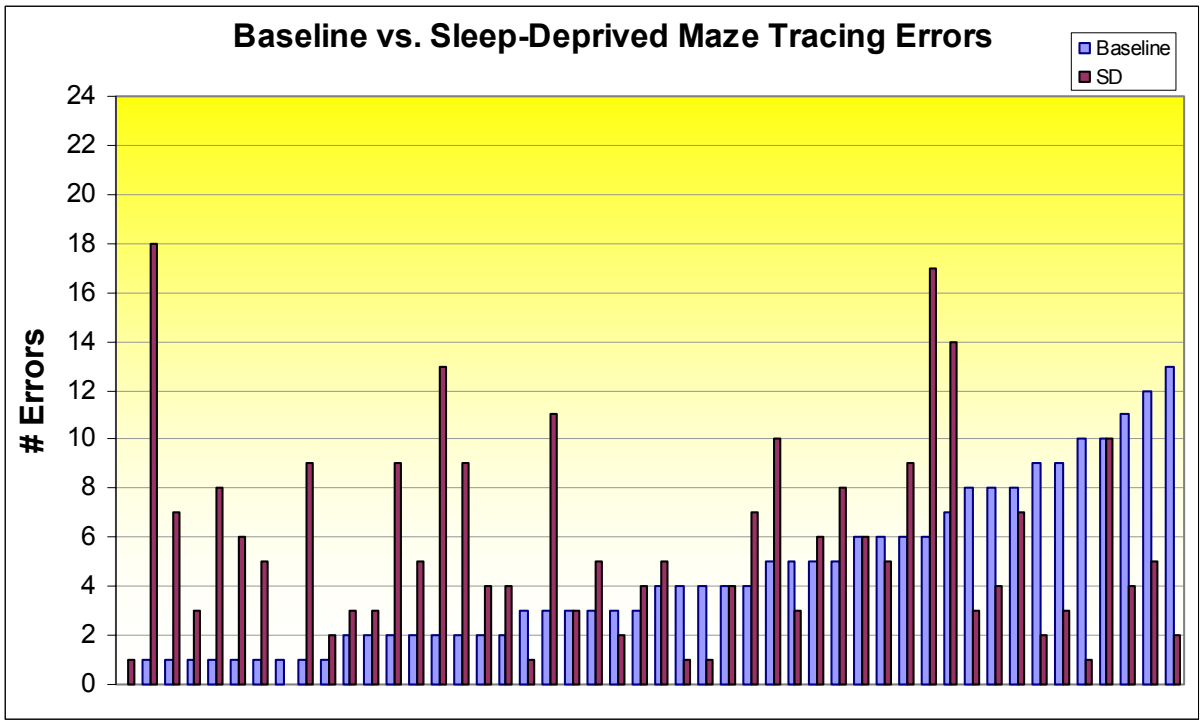


Figure 32. Number of Errors for Maze Tracing Test

The results of the post hoc analysis for MTT errors did not indicate an improvement in performance after SD, contrary to the previous MTT analysis. Rather results indicated a possible trend towards performance degradation. Based on Figure 32, it appears that those participants with few errors during baseline testing had a large increase in errors after SD and that those with eight or more errors during baseline testing were able to reduce errors. Thus, it is possible that maze tracing errors (as defined above) may be a better measure of planning strategy than total number of mazes completed in 3 minutes.

4.6.7 Decision Making (Iowa Gambling Task)

Results of the parametric paired t-test for decision making (as measured by total number of bad decks selected over 100 trials) indicated no degradation in decision making as a result of SD ($p=.214$); however, the nonparametric test approached significance for degraded performance ($p=.059$). The assumptions for a parametric test were not met, and therefore, the nonparametric results were used for discussion purposes. Figure 33 shows the number of selections from risky decks for both baseline and sleep deprived sessions. Both baseline and sleep-deprived data were

normally distributed (p>.15) baseline and sleep deprived sessions had equal variance (p=.890) (See Appendix D).

There was one outlier in the baseline data. The data value (79) was further than two standard deviations from the mean (M =37.9, SD = 15.0); therefore, the outlier data was removed from both sessions. The data subset was then reanalyzed for normality and equal variance (N=36). Reanalysis of the data without outliers indicated that both sessions were normally distributed (p>.15) and had equal variance (p=.411) (see Appendix D).

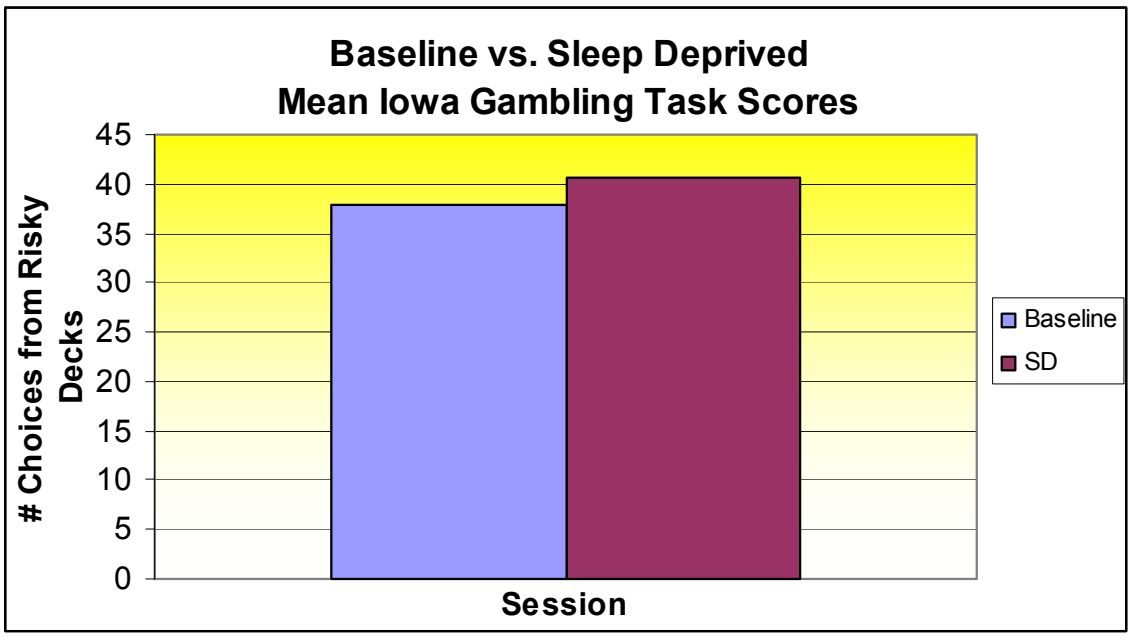


Figure 33. Mean IGT Scores by Session

Figure 34 shows the differences between baseline and sleep deprived scores for each participant. Participants are rank-ordered by baseline score. Eleven participants improved with sleep deprivation, three participants showed no difference in score, and 23 degraded with sleep deprivation. The mean improvement in score was 12.6 (SD=10.8). The mean degradation in score was 10.3 (SD=6.2).

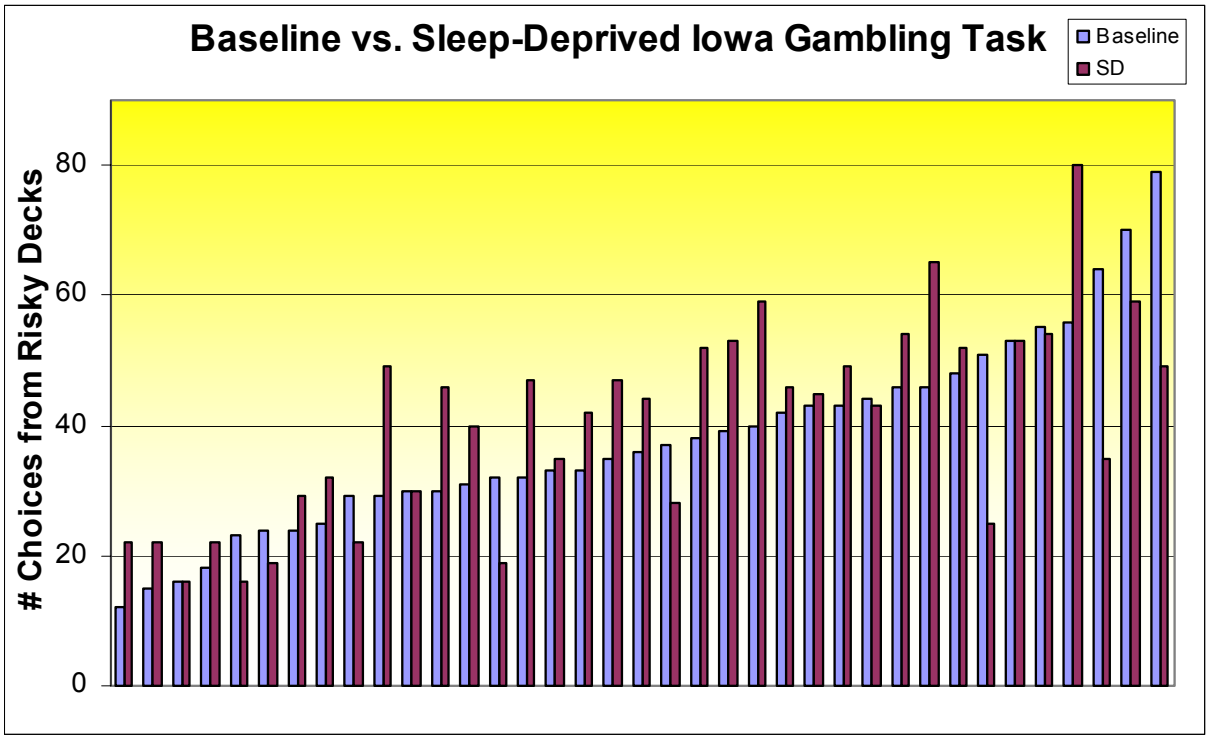


Figure 34. Iowa Gambling Task Performance by Session

The mean baseline score for those participants who improved with sleep deprivation was 46.2 (SD=19.2) while the mean baseline score for those participants who degraded was 34.5 (SD=10.9). The mean baseline score of those who showed no difference was 33.0 (SD=18.7). Participants who improved had significantly higher baseline scores than those who degraded (p=.042). There were no significant differences in baseline scores between those participants whose scores did not change and those who improved (p=.180) or those who degraded (p=.138).

Participants who showed no difference in IGT score had significantly higher WASI IQ scores than those who improved (p=.005) and those who degraded (p=.022). However, when IQ score is covaried for education and language, there was no longer a significant difference in IQ score (See Appendix D).

In the present study, decision-making was measured using total number of trials (out of a possible 100) on which volunteers selected from "bad" (high-risk) decks in the Iowa Gambling Task (IGT). A post hoc analysis to investigate the effects of sleep deprivation on decision making was warranted because the nonparametric paired t-test indicated some possible

degradation in sleep deprivation and results of other studies showed degradation in decision making on the IGT after sleep deprivation (Herscovitch, Stuss et al., 1980) and after frontal lobe damage (Bechara, 1994; Manes, Sahakian et al., 2002). In several previous studies, degraded decision-making was apparent only when the 100 trials of data were further divided into 5 sequential, successive "bins" of 20 trials each. Therefore, a supplemental analysis was conducted in which the 100 trials of data were divided into 5 bins to characterize the performance *pattern* across both rested and sleep-deprived conditions, and to determine whether a trend toward degradation due to sleep deprivation could be seen across the 5 bins.

4.6.7.1 Post Hoc Analysis for Decision Making

Results of the supplementary analysis are shown in Figure 39, in which a net score for each bin was derived by subtracting the number of selections from the bad decks from the number of selections from the "good" (low-risk) decks (thus, positive net scores indicate more selections from the good decks; negative net scores indicate more selections from the bad decks). Bechara et al. (2000) showed that the pattern of deck selection exhibited by their normal (i.e., otherwise healthy, non-sleep deprived) adult control volunteers was characterized by an increasing frequency, across bins, of selection from good decks (i.e., increasingly positive net scores across bins). Figure 35 shows that in the present study, under baseline (rested condition), volunteers performed similarly to the normal controls in the Bechara et al. (2000) study.

In contrast, after 22 hours of sleep deprivation, the pattern of deck selection exhibited by our volunteers across the 5 bins changed – and became more similar to that seen in patients with damage to the ventromedial cortex (Bechara, Tranel, and Damasio, 2000). That pattern was characterized by selection mostly from the good decks for the first few bins followed by a switch to more frequent selections from the high-risk deck for the later bins (resulting in less positive net scores in the later bins).

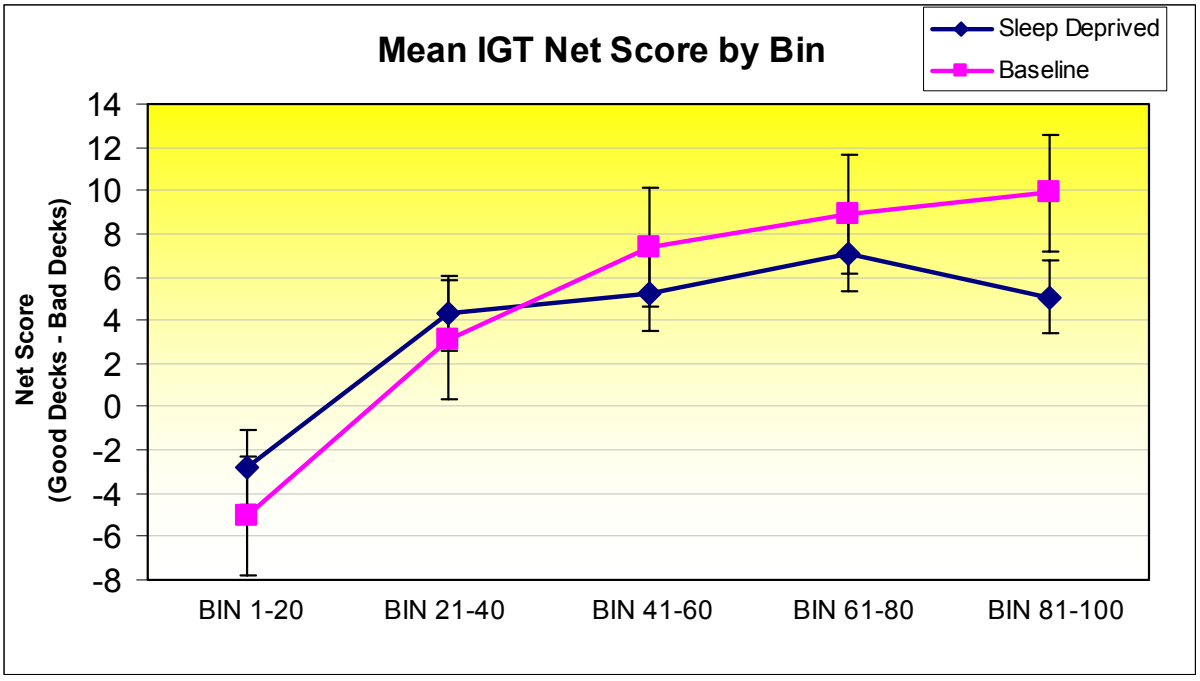


Figure 35. Mean IGT Net Score by Bin

Paired t-tests ($\alpha = .05$) were performed to determine if there was a significant difference in net scores across bins for baseline versus sleep-deprived conditions (see Table 14). Results indicated no statistically significant difference in net scores between rested and sleep-deprived conditions until the third bin, where volunteers' net scores were marginally ($p=0.061$) less positive (indicating more selections from bad decks) than they had been under the rested condition. By the last bin, however, this difference was significant. For the last bin, volunteers' net scores were significantly less positive than they had been under the rested condition ($p=.001$).

Table 14. Paired Differences for IGT by Bin

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Bin 1-20	1.10811	4.28034	.70368	-.31903	2.53524	1.575	36	.124
Bin 21-40	.62162	4.87255	.80104	-1.00297	2.24621	.776	36	.443
Bin 41-60	-1.08108	3.40244	.55936	-2.21551	.05335	-1.933	36	.061
Bin 61-80	-.91892	3.98175	.65460	-2.24650	.40866	-1.404	36	.169
Bin 81-100	-2.40541	4.05832	.66718	-3.75852	-1.05229	-3.605	36	.001

This supplementary analysis indicated that decision-making was in fact degraded by sleep deprivation, but this degradation did not become apparent until later in the IGT task. Although it is not clear why decision-making degradation did not occur until volunteers were well into the IGT task, it may be that fatigue (time on task) interacts with sleep deprivation to unmask cognitive effects. Wilkinson (1965) showed that long-duration (e.g., 30-minute) tasks were more sensitive to sleep deprivation effects than were short-duration (e.g., 1-2 minutes) tasks. In other words, although volunteers can initially perform well when sleep deprived, they may not be able to sustain optimal performance for more than several minutes.

4.6.8 Situation Awareness (WOMBAT)

Situation awareness (as measured by WOMBAT overall score) was not degraded by sleep deprivation (p=.257). Figure 36 shows the mean overall WOMBAT scores for both baseline and sleep deprived sessions. The mean baseline score was 94.3 (SD=50.1) and the mean SD score was 98.9 (SD=55.5).

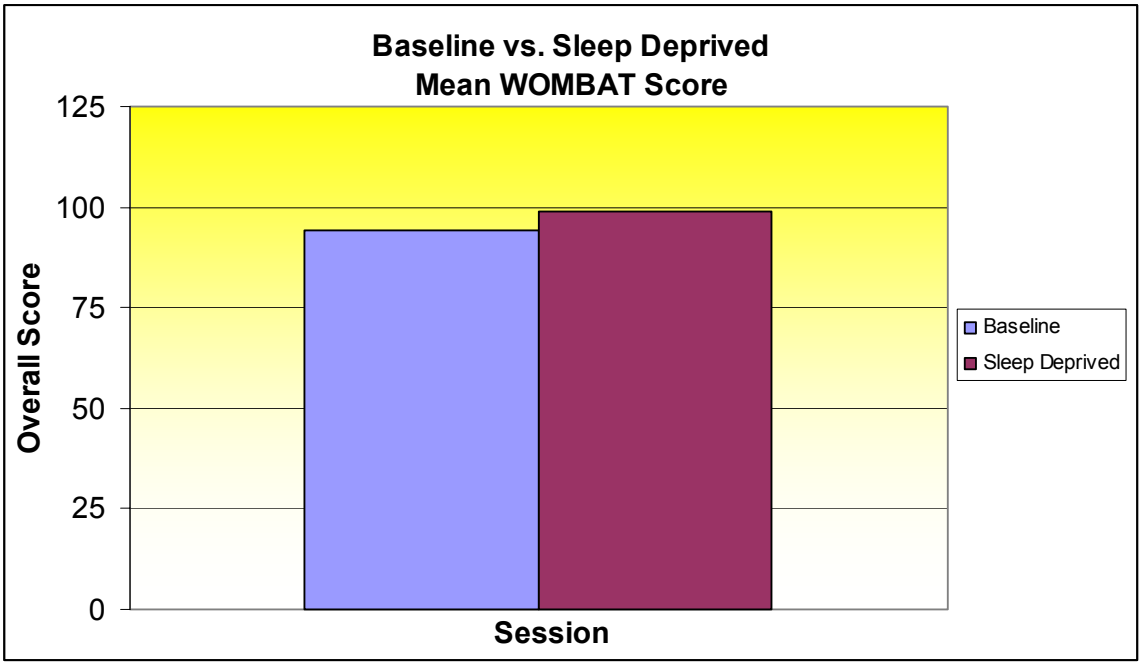


Figure 36. Mean WOMBAT Scores

Figure 37 shows the differences between baseline and sleep deprived scores for each participant. The baseline data for WOMBAT Overall Score followed a normal distribution ($p>.15$). Data from sleep-deprived subjects was also normally distributed ($p>.15$). Baseline and sleep deprived sessions had equal variance ($p=.492$) (See Appendix D). There were no outliers.

Twenty-two participants improved from baseline score and 26 degraded from baseline score. The mean improvement in score was 23.6 points (SD=20.4) while the mean degradation was 17.9 points (SD=15.9). The mean baseline score for those participants who improved with sleep deprivation was 103.2 (SD=49.1) while the mean baseline score for those participants who degraded was 86.8 (SD=50.8). There was no significant difference between baseline scores of those who improved vs. those who degraded with sleep deprivation ($p=.132$).

Results of an ANOVA ($\alpha = .05$) indicated there was no significant difference in age or IQ score between those who improved and those who degraded. The CMH measure of general association approached significance ($p=.073$), indicating a potentially significant association between improvement vs. degradation and morningness-eveningness ($p=.008$) (See Appendix D).

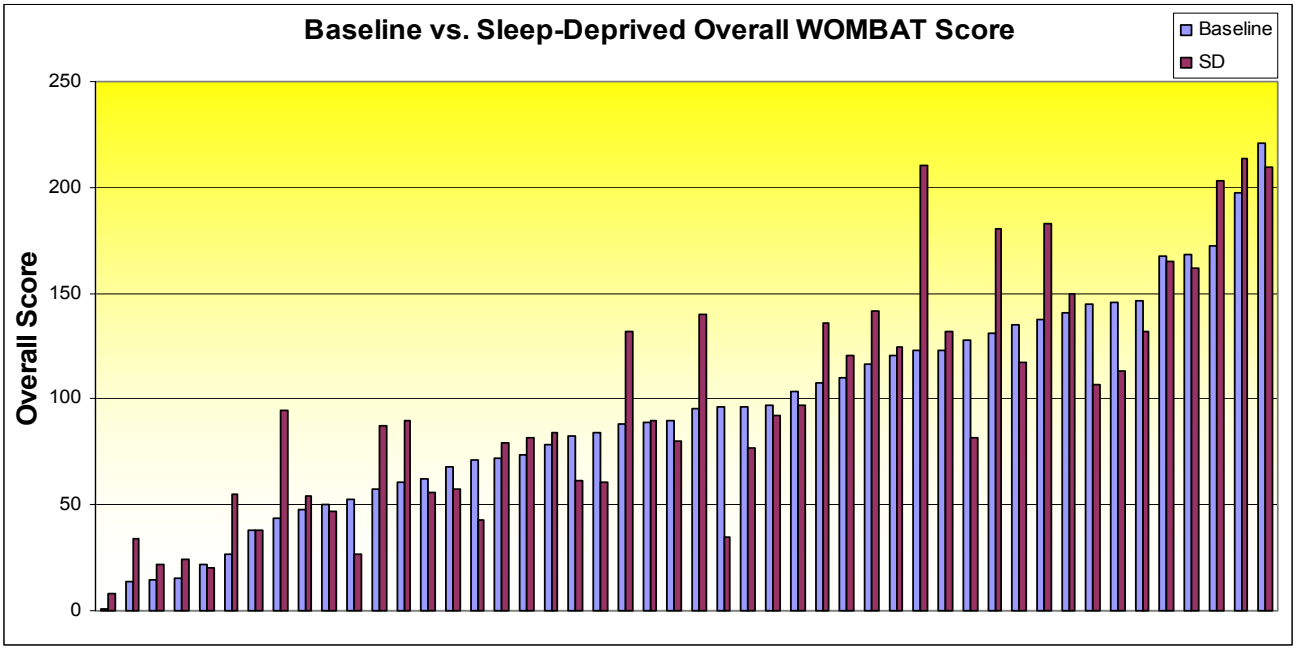


Figure 37. Overall WOMBAT Scores by Session

Because the paired t-test for WOMBAT overall score did not yield significant results, a post hoc analysis on additional WOMBAT dependent variables automatically collected by the software program was conducted using paired t-tests.

4.6.8.1 Post Hoc Analysis for Situation Awareness

Several additional performance variables were calculated by the WOMBAT program. Variables selected for analysis and their respective definitions are as follows:

1. Total Tracking—sum of the product of the tracking performance achieved at each point in the test and the tracking worth at the time increment
2. Overall Tracking Percent—100 x (Tracking score / perfect tracking score)
3. Total Collision Detection—score earned by predicting collisions of the targets on the grid.
4. Total Solid—Figure Rotation score x Figure Rotation worth
5. Total Quadrant—Quadrant Location Score x Quadrant Location worth
6. Total Sequences Mastered—number of sequences mastered in the Quadrant-Location task

7. Total Two Back—Digit Canceling score x Digit Canceling worth

8. Total Bonus—sum of Figure Rotation score, Quadrant Location score, And Digit Canceling score

9. Missing Targets Found—total number of missing targets found

10. Time Missing Before Found—time elapsed between activating the grid to search for a target and finding the target

Table 15 lists the results of the normality of paired differences analyses for the additional WOMBAT dependent variables. Assumptions were not met for the Overall Tracking Percent or the Visible Collisions (See Appendix D).

Table 15. Paired T-test Assumptions for Additional WOMBAT Variables

Dependent Variable	Normality of Paired Differences (p-value)
Total Tracking	Normal (p=.146)
Overall Tracking %	Not Normal (p=.032)
Total Collision Detection	Normal (p>.15)
Total Solid	Normal (p=.049)
Total Quadrant	Normal (p>.15)
Total Sequences Mastered	Normal (p>.15)
Total Two Back	Normal (p>.15)
Total Bonus	Normal (p>.15)
Missing Targets Found	Normal (p>.15)
Time Missing Before Found	Normal (p>.15)
Visible Collisions	Not Normal (p<.010)

Table 16 lists results of paired t-tests for each of the post hoc dependent variables. Significant differences in baseline vs. sleep deprived performance were found for the following variables: Total Solid (p=.004), Total Quadrant (p=.033), and Total Bonus (p=.020). Results indicated that performance on these measures improved with sleep deprivation, which is consistent with the apparent overall maintenance of performance on the WOMBAT under sleep deprived conditions. Thus, it did not appear that any measure on the WOMBAT task was degraded by sleep deprivation (regardless of whether the dependent measure indexed SA).

Results of the present study did not support the hypothesis that situation awareness would be degraded by sleep deprivation. Several authors have hypothesized that SA would be degraded by sleep deprivation and provided anecdotal evidence of degraded SA based on errors and accidents (Steinweg, 1995; Jones and Endsley, 1996; Jay, 2003); however, anecdotal

evidence of degraded SA was not supported by findings from the present study using the WOMBAT.

Table 16. Paired T-tests Results for Post Hoc WOMBAT Dependent Variables

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Total Tracking	-.2875	21.876	3.1576	-6.6398	6.0648	-.091	47	.928
Overall Tracking Percent	.3167	6.8992	.9958	-1.6866	2.3200	.318	47	.752
Total Collision Detection	.4625	8.9951	1.2983	-2.1494	3.0744	.356	47	.723
Total Solid (Figure Rotation)	-3.0833	6.9677	1.0057	-5.1066	-1.0601	-3.066	47	.004*
Total Quadrant	-1.6021	5.0641	.7309	-3.0726	-.1316	-2.192	47	.033*
Total Sequences Mastered	-.250	1.042	.150	-.552	.052	-1.663	47	.103
Total Two Back	-.0854	5.2363	.7558	-1.6059	1.4350	-.113	47	.910
Total Bonus	-4.7396	13.635	1.9674	-8.6975	-.7817	-2.409	47	.020*
Number Missing Targets Found	-.104	3.855	.556	-1.224	1.015	-.187	47	.852
Time Missing Before Found	-.079203	1.9926	.28760	-.65779	.499387	-.275	47	.784
Visible Collisions	.396	6.105	.881	-1.377	2.168	.449	47	.655

4.7 The Relationship Between Self-reported Sleepiness And Complex Task Performance Requiring SA.

This dissertation proposed to investigate the relationship between perceived workload and complex task performance requiring SA under sleep deprived conditions; however, the NASA-TLX subjective workload assessment was not administered to participants following WOMBAT testing and thus, an analysis of the relationship between subjective workload and complex task performance was not conducted.

An alternative objective analysis of workload is to measure performance on secondary tasks (Sanders and McCormick, 1993). Therefore, a supplemental workload analysis was conducted based on secondary task measures collected by the WOMBAT software. The bonus tasks in WOMBAT are secondary to the tracking task, and thus an analysis of performance on secondary (i.e., bonus) tasks may yield some insight into the relationship between workload and

performance on a complex task requiring SA. Results of the WOMBAT analysis indicated that neither the primary tracking task performance nor total bonus task performance was degraded after sleep deprivation. In fact, performance on the secondary quadrant location and figure rotation (i.e., total solid) tasks actually improved. Therefore, an objective measure of workload indicated that workload did not increase with sleep deprivation. A correlation between an objective workload measure (i.e., secondary task performance) and subjective workload ratings was not possible because the NASA-TLX was not administered; however, future research might yield interesting results with regard to the correlation between objective and subjective workload under sleep deprived conditions. Because subjective workload was not assessed, the relationship between complex task performance requiring SA and perceived sleepiness was analyzed post-hoc as an indication of the relationship between subjective perception and complex task performance.

To test the hypothesis that there was a significant relationship between self-reported sleepiness and complex task performance requiring SA, a Pearson correlation analysis (α=.05) was conducted between WOMBAT overall score and Stanford Sleepiness Scale ratings, and between WOMBAT overall score and Karolinska Sleepiness Scale ratings. Results of several studies have shown that there is no effect on Type I and Type II error when ordinal variables such as those derived from the Stanford and Karolinska Scales are used in interval data analysis techniques and using the Pearson r with ordinal data is now the norm in social science (Garson, 2005). Tables 17 and 18 list the correlation matrices for baseline and sleep deprived data, respectively.

Table 17. Correlation Matrix for Baseline Data

	Karolinska	Stanford
Stanford	0.723	
	0.000*	
WOMBAT	-0.139	0.040
	0.351	0.787

Cell Contents: Pearson correlation
P-Value

Table 18. Correlation Matrix for Sleep Deprived Data

	Karolinska	Stanford
Stanford	0.880	
	0.000*	
WOMBAT	-0.027	-0.101
	0.856	0.500

Cell Contents: Pearson correlation
P-Value

Results of the correlation analysis did not support the hypothesis that self-reported sleepiness is significantly related to complex task performance during either baseline or sleep deprived testing sessions. Because sleepiness was not monitored periodically throughout the 60-minute WOMBAT test session, it was not possible to use regression to predict complex task performance as a function of sleepiness. These results contrast with those from the PVT analysis in which self-reported sleepiness significantly predicted mean RT, indicating yet another discrepancy between the effects of sleep deprivation on simple tasks versus executive function-type tasks.

4.8 The Effects of SD on Complex Task Performance requiring SA over Time

Simple linear regression was conducted using WOMBAT interval score data to determine the effects of sleep deprivation on complex task performance over time. Figure 38 shows mean overall WOMBAT scores in five-minute increments for both baseline and sleep-deprived testing sessions.

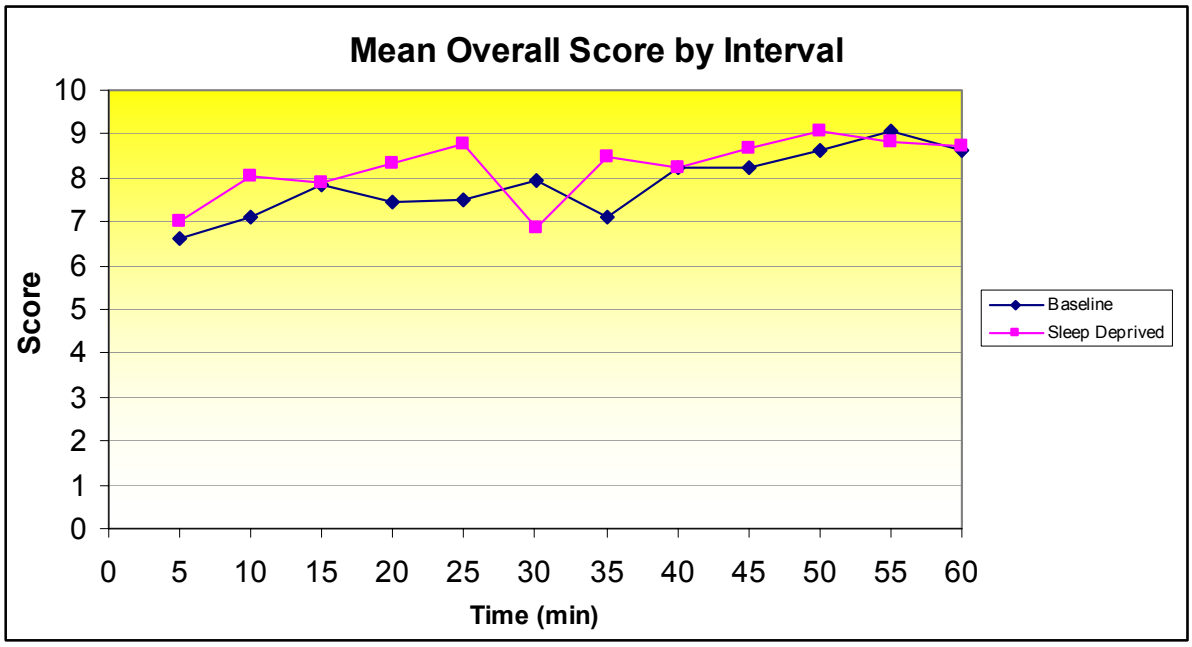


Figure 38. Mean WOMBAT scores over time

Complex task performance (as measured by overall WOMBAT scores) improved over the 60-minute test interval during the baseline session. The baseline linear regression equation was as follows: Baseline Mean Interval Overall = 6.70 + 0.0356 Time (min). Table 19 lists results of the ANOVA for the WOMBAT regression. The linear model accounted for 73.3% of the variance (p=.000) and the residuals were normally distributed (p>.15)(see Appendix H). . Figure 39 shows the linear fit for the baseline regression.

Table 19. ANOVA for Baseline WOMBAT Regression

Source	DF	SS	MS	F	P
Regression	1	4.5301	4.5301	31.19	0.000
Residual Error	10	1.4526	0.1453		
Total	11	5.9827			

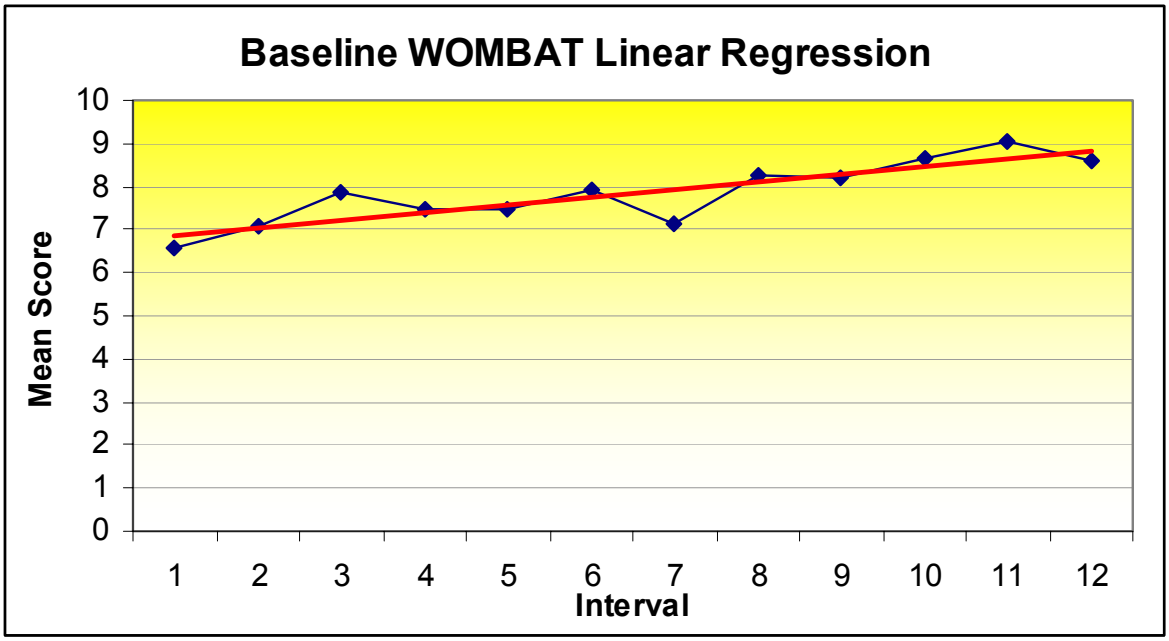


Figure 39. Regression plot for Baseline WOMBAT Interval Scores

Figure 40 shows the linear fit for the sleep deprived WOMBAT regression. Similar to the baseline condition, the regression indicated that performance improved over the 60-minute test session following sleep deprivation. The regression equation for the sleep deprivation condition was as follows: SD Mean Interval Overall Scores = 7.43 + 0.0249 Time (min). Table 20 shows the ANOVA for the sleep deprived WOMBAT Regression.

Table 20. ANOVA for Sleep Deprived WOMBAT Regression

Source	DF	SS	MS	F	P
Regression	1	2.2094	2.2094	7.15	0.023
Residual Error	10	3.0902	0.3090		
Total	11	5.2996			

The regression equation was significant (p=.023); however, the equation only accounted for 35.9% of the variance (see Appendix H). The regression residuals were normally distributed (p>15) and a linear fit was appropriate based on the plots of residuals vs. fits and residuals vs. order of the data (see Appendix H).

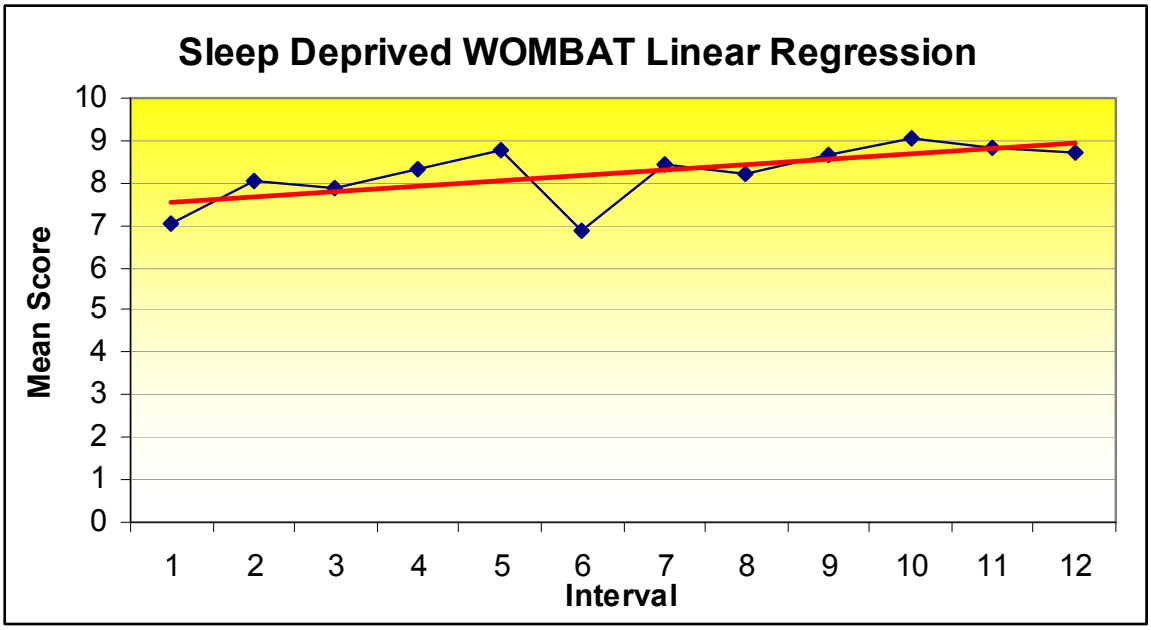


Figure 40. Regression plot for Sleep Deprived WOMBAT Interval Scores

To determine whether there was a difference in the slope of improvement over time between baseline and sleep deprivation conditions, an F-test was conducted on the slopes of the two regression lines. Results indicated that the slopes did not differ (p=.352). The intercept for the sleep deprivation condition was larger than that for the baseline condition and the difference was marginally significantly (p=.063). This suggests a possible learning effect because the first interval score for the sleep deprived testing session is higher than the first baseline interval score. Thus, WOMBAT performance improved over time under both baseline and sleep deprivation conditions and it appears that participants may have had improved performance at the start of the second testing session.

A two-factor ANOVA (α=.05) was also conducted to determine if there was a significant difference in WOMBAT score between sessions (baseline vs. sleep deprived) or between intervals (1-12). The sphericity assumption was not met; therefore, a Geisser-Greenhouse correction was used. Table 21 shows the results of the ANOVA.

Table 21. Two-factor WOMBAT ANOVA

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
session	Sphericity Assumed	42.014	1	42.014	1.308	.259
	Greenhouse-Geisser	42.014	1.000	42.014	1.308	.259
Error(session)	Sphericity Assumed	1510.208	47	32.132		
	Greenhouse-Geisser	1510.208	47.000	32.132		
interval	Sphericity Assumed	418.307	11	38.028	4.491	.000
	Greenhouse-Geisser	418.307	5.116	81.770	4.491	.001
Error(interval)	Sphericity Assumed	4377.763	517	8.468		
	Greenhouse-Geisser	4377.763	240.436	18.208		
session * interval	Sphericity Assumed	123.246	11	11.204	1.702	.070
	Greenhouse-Geisser	123.246	8.119	15.179	1.702	.095
Error(session*interval)	Sphericity Assumed	3404.182	517	6.584		
	Greenhouse-Geisser	3404.182	381.614	8.920		

Results indicated a significant main effect for interval (p=.001). There was no significant difference across session (p=.259), and the Session x Interval interaction approached significance (p=.095). To isolate significant differences for the interval main effect, post hoc paired comparisons among means were conducted using Tukey's HSD test with α=.05 (See Appendix X). Results indicated that WOMBAT scores during Interval 1 were significantly lower than scores during Intervals 10 and 11. The Tukey HSD post hoc results were compared to Bonferroni-corrected paired comparisons (α=.05) to determine if a more robust test yielded different results (See Appendix H). Bonferroni results yielded the following significant paired comparisons:

- Interval 1 < Intervals 10 and 11
- Interval 6 < Intervals 10 and 11

Because of the exploratory nature of this analysis, it was concluded that a less conservative approach was warranted and the Bonferroni results (which yielded additional significant differences) were used for discussion purposes. Based on the regression and ANOVA analyses, it was concluded that WOMBAT performance improved over time. This conclusion was further supported by the ANOVA analysis which indicated a trend toward significantly improved performance over time.

Because vigilance on simple reaction time tasks is degraded as a function of time on task (Wesensten et al., 2004), it was expected that WOMBAT performance also would degrade across the 60-minute test session. However, results did not support this hypothesis. In fact, WOMBAT

performance significantly improved across the 60-minute session during both baseline and sleep deprived conditions. This within-session improvement suggests that the WOMBAT is subject to learning effects.

4.9 Discussion of Executive Function Results

In this dissertation, it was hypothesized that situation awareness, planning, and decision making would be degraded by sleep deprivation (SD). This hypothesis was based on results of previous studies in which decrements in executive functioning (i.e., innovative, flexible thinking and strategic planning) were seen after one night of sleep loss (Harrison and Horne, 1999; Wimmer, Hoffman et al. 1992; Bandaret, Stokes et al., 1981). It was also hypothesized that working memory, attention, and vigilance would be degraded by sleep deprivation based on results from the Walter Reed laboratory as well as other publications (Himarashee, Banerjee et al., 2002; Harrison and Horne, 1999; Wimmer, Hoffmann et al., 1992; Horne, 1988; Bandaret, Stokes et al., 1981; Thomas, Sing et al., 2000). Results from paired t-tests failed to confirm degradation on any executive function task as a result of sleep deprivation. Likewise, working memory performance was not degraded by sleep deprivation. The results did show degradation in attention as measured by psychomotor vigilance reaction time. The following sections describe several possible explanations for these results.

4.9.1 Effects of SD on Cognitive Tasks – Compensatory Brain Activation

One explanation for the nonsignificant findings with regard to tasks of executive function has been proposed by Drummond et al. (2000). The author hypothesized that performance on some tasks may be maintained during sleep deprivation because the brain somehow compensates for sleep deprivation-induced relative deactivation in certain areas by increasing relative activation in other areas. Specifically, Drummond et al. (2000) proposed that the intensity of compensatory activation is dependent upon the specific cognitive demands of the task (with more complex tasks resulting in greater compensatory activation than simple tasks). Further, Drummond et al. (2000) proposed that tasks which under rested conditions preferentially activate the prefrontal cortical areas of the brain (an operational definition of "task complexity") result in greater relative compensatory activation during SD than tasks which do not rely on the prefrontal cortex. Thus, Drummond et al's hypothesis might better be referred to as the "brain region-specific" hypothesis.

(Killgore, Balkin, and Wesensten, in submission) further proposed that compensation is differentially activated by two subregions of the PFC (the dorsolateral and ventromedial regions) in response to information processing demands. They hypothesized that the executive processes of lateral and dorsolateral prefrontal cortex are consciously available and can actively recruit compensatory processing. Conversely, the processes of the ventromedial prefrontal cortex are automatic and subconscious, and thus, less able to be actively recruit compensation.

This dissertation did not directly test the hypothesis that compensatory brain activation resulted in sustained executive function performance, and therefore, no definitive statements can be made as to whether the hypothesis accounts for the specific results of this study. Furthermore, research has not indicated whether the relative activation seen in response to sleep deprivation is caused by individual differences in response to sleep deprivation or whether the response is universally task-specific. The findings of previous studies that relate directly to the cognitive processes tested in this study are presented only to show the potential plausible explanation for this study's results and to identify an area of future research that may elucidate these results further. To hypothesize which executive function tasks may differentially activate compensatory information processing mechanisms, additional brain imaging studies are necessary to determine the specific activation patterns for the executive function tasks used in this study.

Results of brain imaging studies lend support to the cognitive demand specific hypothesis on several cognitive tasks that could possibly be extrapolated to explain the findings in this study. For example, increased relative cerebral activation in response to SD has been found during verbal learning (Drummond et al., 2000) and divided attention tasks (Drummond et al., 2001), both of which rely heavily on the prefrontal cortical areas of the brain. In contrast, decreased relative activation was found during a serial subtraction task (Drummond et al, 1999) and a novelty processing task (Gosselin, DeKoninck, and Campbell, 2005), both of which rely less heavily on the prefrontal cortex. In other words, during sleep deprivation, increased relative activation was seen for tasks that rely heavily on frontal areas of the brain whereas decreased relative activation was seen for those tasks which did not rely on frontal areas of the brain. Importantly, compensatory brain activation resulted in sustained performance on the verbal learning task (Drummond and Brown, 2001), but did not sustain performance on the serial

subtraction task, novelty processing task, or a divided attention task (Drummond et al., 2001). Thus, there are conflicting results showing increased activation with both maintenance and degradation of performance during sleep deprivation.

Results from imaging studies of a variety of working memory tasks indicate that some brain areas are consistently activated by these types of tasks, including the dorsolateral PFC (D'Esposito et al., 1995). Also, results from a PET study by Haut et al. (2000) indicated relative activation of the PFC during a letter-number sequencing task similar to that used in the present study. Interestingly, there is also some evidence that, within a given task, increased relative activation is associated with better performance during sleep deprivation. Chee and Choo (2004) conducted an fMRI study to determine the neurobehavioral effects of sleep deprivation on two working memory tasks that required maintenance of verbal working memory and one that required manipulation of items in verbal working memory. They found that overall, response time was degraded by sleep deprivation. However, in the same study, greater relative parietal lobe activation was correlated with less impaired word recall; and reduced relative deactivation of midline frontal regions during more complex tasks was associated with better performance. Chee and Choo (2004) hypothesized that degradation in behavioral performance was the result of a reduction in cognitive resources available for recruitment (as indexed by decreased relative cortical activation). Thus, compensatory activation may have contributed to sustained performance on the working memory task (LNS) in this study.

With respect to planning, Peterson, van Mier, Fiez, and Raichle (1998) found that different areas of the brain were activated during unpracticed maze tracing (namely, the right premotor and parietal cortex and left cerebellar hemisphere) compared to practiced maze tracing, which activated the medial frontal cortex or "supplementary motor area." Thus, despite its potential for measuring the construct of planning, the present results suggest that volunteers must obtain substantial practice on this task prior to the study and/or that the task is unsuitable for repeated-measures designs. Because of the learning effect, it is not clear whether this task is actually sensitive to SD. Also, it is not clear whether the motor and cognitive components of this task can be separated out – and only the cognitive component would have been of interest in the present study.

The results on the decision making (IGT) task appear to be inconsistent with the Drummond et al.'s (2000) hypothesis. That is, results of imaging studies have shown substantial activation of the PFC during performance of decision-making tasks in rested individuals (Ernst et al., 2002; Bechara et al., 1999). Thus, as per Drummond et al. (2000), greater compensatory activation during SD should have been expected on the decision-making task, resulting in sustained performance. However, the present results indicated an impairment in decision-making with SD. One possible explanation for these apparently inconsistent results is the time-on-task or fatigue effect referred to above. That is, under SD, volunteers actually did perform similarly to rested conditions for the first two bins. It was not until the third bin that performance began to deteriorate. Secondly, the implicit learning component of the IGT may not be available to conscious processing and degraded performance could therefore be attributed to the differential compensatory response hypothesized by Killgore et al. (in submission). Neuroimaging of the IGT under SD conditions would yield better insight into the nature of compensatory response over time.

With respect to reasoning tasks (e.g., LS and NS), Horne (2000) posited that rule-based or logical reasoning tasks are more resilient to sleep deprivation than other cognitive tasks because critical reasoning is not solely dependent upon the PFC. Results of an imaging study by Drummond et al. (2004) indicated that after 35 hours of SD, performance of a logical reasoning task resulted in increased relative activation of the bilateral inferior parietal lobes, bilateral temporal cortex, and left inferior and dorsolateral prefrontal cortex. Importantly, Drummond et al. (2004) also reported that performance on this task was not degraded by SD. An fMRI study by Goel and Dolan (2004) showed relative activation of the left medial frontal gyrus, the left cingulate gyrus, and the left superior frontal gyrus during an inductive reasoning task. A follow-up study by Goel and Dolan (2004) found relative activation of left lateral prefrontal and bilateral dorsal frontal, parietal, and occipital cortices for both inductive and deductive reasoning task; however, performance of an inductive reasoning task showed greater relative activation of the left dorsolateral prefrontal gyrus than deductive reasoning task. Patterns of activation for both deductive and inductive reasoning tasks are primarily located within the PFC and it is possible that participants in this study were able to recruit additional cognitive resources on the inductive reasoning similarly to the deductive reasoning task.

The investigation of complex task performance requiring SA (WOMBAT) also showed resilience to the effects of SD. Drummond et al. (2004) posited that these results may, in fact, be due to the facilitation of the compensatory response as a result of task difficulty. Thus, in the present dissertation, the ability to sustain performance on WOMBAT may have been a function of task difficulty. To date, no studies have been published that describe the pattern of brain activation associated with WOMBAT performance. Results from a neuroimaging study comparing relative brain activation for complex versus simple tasks indicated that performance was maintained after sleep deprivation for complex tasks but not simple ones (Chee and Choo, 2004). The results of this study corroborate those of Chee and Choo (2004). In the present study, simple reaction time for the PVT was degraded by sleep deprivation whereas performance on WOMBAT was not. Furthermore, there was no evidence for degradation of vigilance during WOMBAT performance – in fact, vigilance performance improved slightly over the 60-minute WOMBAT test session. Therefore, in this dissertation it is posited that performance on WOMBAT was sustained during SD because it was a complex task.

In sum, the compensatory brain activation hypothesis provides a possible explanation for sustained performance during sleep deprivation on some of the executive function tasks used in this study. To draw firm conclusions about the neurobehavioral mechanisms by which performance was sustained after SD, brain imaging studies are necessary to specify patterns of brain activation, most notably relative activation of the PFC. Secondly, any potential practice or implicit learning effects must be removed from the tasks such that relative brain activation reflects compensation rather than practice or learning.

4.9.2 Individual Differences in Tolerance to Sleep Deprivation based on demographic data

The goal of demographic data analyses was to identify individual difference factors that may account for differential sensitivity to sleep deprivation (SD). The general effects of sleep deprivation on performance are well known; however, how much variability exists with regard to vulnerability to sleep deprivation is not known (Van Dongen, Baynard, Maislin, and Dinges, 2004; Harma, 1995). While performance and subjective sleepiness in some individuals degrades rapidly and substantially during sleep loss, performance and subjective sleepiness in other individuals appears to be "resilient" to sleep loss. Furthermore, this resiliency appears to

be a stable, trait-like characteristic of individuals rather than a result of recent sleep history (Van Dongen et al., 2004).

Variability in individual tolerance to sleep deprivation is potentially attributable to a variety of physiological, psychological, and environmental factors. To date, however, few papers have been published in which factors that may influence individual variability in tolerance to sleep deprivation have been examined. Factors which have been examined include age, gender, sleep flexibility, personality, and physical fitness (Harma, 1995). However, none of these factors accounts for a substantial amount of variability in performance, and as Akerstedt (1999) and Kerkhof (1985) report, results have been inconclusive and contradictory regarding the correlation between these factors and an individual's ability to tolerate sleep loss. Although Giam (1997) suggested that physically active individuals require more sleep to recover, Harma (1995) suggests that physically fit individuals actually report less sleepiness during night shifts. Studies on the effects of personality have produced mixed results as well (as cited in Friedmann et al., 1977).

One factor that has received some attention is "chronotype" or morningness-eveningness preference (Akerstedt, 1991). In the present study, morningness-eveningness was significantly correlated with sleepiness ratings and predicted performance on two cognitive tasks—deductive reasoning (NS) and WOMBAT. There was a significant general association between morningness-eveningness and improvement vs. degradation on the deductive reasoning task (NS) and WOMBAT. Additionally, there was a significant general association between occupation and improvement vs. degradation on the NS task as well as a significant general association between ethnicity and improvement vs. degradation on the MTT task. Other than these latter findings, no other demographic factors predicted vulnerability to SD. Age did not affect improvement or degradation in this study; however, the age range in this study was limited (18-33 years). Results from several sleep deprivation studies have shown that older participants are more resilient to the behavioral effects of sleep deprivation than younger participants (as cited in Philip et al., 2004). Thus, the restricted age range in this study may account for the lack of significant age-related findings. In sum, the results of this study did not point to a universally applicable demographic factor or factors that significantly predict resilience to SD.

4.9.3 Power of Paired T-tests

Although the cognitive demand specific hypothesis is one possible explanation for the results attained in this study, some consideration must be given to the issue of statistical power as well. Table 22 lists the observed power of each paired t-test both with and without outliers (See Appendix F for detailed results). When outliers were included, observed power for the PVT and MTT were high (1.0 and .99, respectively). The observed power for the LS, IGT, WOMBAT, NS, and LNS (.18, .24, .20, .19, .05, respectively) was low and suggested inadequate power for these tasks. When outliers were excluded from the data, observed power for both the PVT and MTT (1.0 and .82, respectively) were within acceptable limits for adequate statistical power. Observed power for LS and IGT (.50 and .44, respectively) was lower. However, observed power for WOMBAT, NS, and LNS (.21, .12, and .07, respectively) was low and suggested inadequate statistical power for these tasks. Thus, when outliers were included in the data, the observed power of the MTT and NS improved; the observed power of PVT and WOMBAT was maintained; and the observed power of LS, IGT, and LNS was reduced.

Table 22. Power of Paired T-tests

Cognitive Test	Power (no outliers)	Power (with outliers)
PVT	1.0	1.0
MTT	.82	.99
LS	.50	.18
IGT	.44	.24
WOMBAT	.20	.20
NS	.12	.12
LNS	.07	.05

Because the power of the paired t-tests was moderate to low for the majority of the cognitive tasks, the results must be interpreted cautiously. One factor that may have reduced statistical power was the substantial between-subjects performance variability on the cognitive tests. This variability was seen for both baseline and sleep deprived sessions. To compensate for the low power, more robust post hoc analyses were conducted in addition to the Tukey post-hoc comparisons. Results of these additional post hoc analyses suggested (1) that decision making is indeed degraded by SD as a function of time on task (as indicated by significantly fewer selections from good decks towards the end of the testing session (i.e., 81^{st}-100^{th} bin)) and (2)

that planning errors may increase with SD (as indicated by a trend towards more maze tracing errors). Inclusion of these more robust post-hoc analyses revealed effects of SD on executive function that may have gone undiscovered had only paired t-tests been used.

4.9.4 Limitations of Selected Cognitive Tasks and Amount of Sleep Loss

Two additional factors that may have impacted the present results are (1) amount of sleep loss accrued at testing time and (2) limitations of the executive function tasks. Executive function tasks were selected based on one or more of the following criteria: (1) face validity (WOMBAT); (2) previously demonstrated sensitivity to sleep loss (LS, NS); and/or (3) previously demonstrated sensitivity to PFC damage (LNS, IGT, MTT).

4.9.4.1 Amount of Sleep Loss

Even though some of the tasks used in the present study had previously been demonstrated to be sensitive to 2+ nights of sleep loss, it is possible that these tests were not sensitive to only one night of SD. Sleep loss effects were seen on the PVT, a finding which suggests that 28 hours of sleep deprivation was sufficient to cause decrements on at least some aspects of cognitive performance. In addition, it was possible that volunteers in the present study accumulated some sleep during the sleep deprivation period. Maintenance of wakefulness tests (MWT) were administered every two hours throughout the sleep deprivation period, and during the MWT, participants could have accumulated up to 15 minutes of sleep. However, due to equipment failure, much of the sleep/wake data collected during the MWTs was lost; therefore, it is not known how much actual sleep volunteers accrued during the MWTs.

4.9.4.2 Task Factors

Task factors to consider include (1) dependent variables analyzed; (2) software considerations (WOMBAT); (3) equivalence of different forms of the same test; and (4) practice/learning/order effects.

4.9.4.2.1 Dependent variables analyzed

Additional dependent variables were analyzed for LS (inductive reasoning) and NS (deductive reasoning) tasks (i.e., net score—number of correct responses minus number of incorrect responses) as well as the MTT (planning) task (i.e., number of errors). However, results for these dependent variables were not different than those from the originally selected dependent variables (which included number correct for LS and NS, and number of mazes completed in 3 minutes for MTT). None of the dependent variables available for analysis for

LS, NS, or MTT included measures of reaction time (speed). Results from numerous previous studies of sleep deprivation have shown that reaction time measures are the most sensitive to sleep deprivation whereas measures of accuracy (e.g., net score, errors) do not appear to be affected by sleep loss. Such findings have led to speculation that volunteers who are sleep deprived sacrifice speed to maintain accuracy (Balkin et al., 2000), although the degree to which this trade-off is under conscious control has not been established. Nonetheless, such results do suggest that the dependent measures available for analysis will impact whether the task is sensitive to sleep loss -- and it may simply be that for dependent variables such as number correct/errors, the range is often too small to detect sleep deprivation effects (particularly if the task is not difficult).

4.9.4.2.2 Software considerations (WOMBAT)

The WOMBAT scoring algorithms did not capture small deviations in performance despite an apparently adequate scoring range (approximately 200 points). This failure to capture small deviations may have been a function of other aspects of the WOMBAT software, (e.g., the observation that WOMBAT task difficulty did not appear to be updated rapidly in response to improved performance). As a result, it appeared that participants generally were not challenged to the point of reaching task overload. Such a delayed software response also implies that frequent (but perhaps small) deviations in participants' performance either were not captured or were not reported as part of the WOMBAT scoring output (only gross performance scores such as overall score, total bonus, etc. were available for analysis). Adjustments to the WOMBAT software to correct for these deficits might improve its sensitivity to sleep deprivation.

4.9.4.2.3 Equivalence across test forms.

Although it is possible that the failure to find stronger sleep deprivation effects may have been due to a lack of equivalence among different versions of a particular test form, in the present study it was presumed that equivalence would not be an issue for several reasons. First, publishers' claims for the LS, NS, MTT, IGT were that the various forms of the test had been evaluated and been determined to be equivalent (French, Ekstrom et al., 1976; Bechara, 1994). Second, LNS test forms were created using a random number/letter generator in Excel. Additionally, tests forms were counterbalanced across participants for baseline and sleep deprivation sessions in order to control for order effects. It is possible, however, that the WOMBAT was not equivalent across sessions: although initial WOMBAT start-up conditions

were randomly generated at the beginning of each test, WOMBAT was purposely designed to adapt to each individual user's performance. Thus, by definition each WOMBAT testing session would be unique (and perhaps nonequivalent).

4.9.4.2.4 Practice/learning/order effects

Results for the MTT (planning) indicated that performance actually improved from baseline to SD which suggested a strong learning component for this task. Results from a previous sleep restriction study conducted at the Walter Reed laboratory (Balkin et al., 2000) indicated substantial practice effects even for tasks considered to be relatively simple and easy to learn (e.g., a serial addition and subtraction task; a 4-choice reaction time task). In the present study, learning effects could have masked the effects of sleep deprivation. Learning effects on some tasks, however, were unavoidable. For example, only two versions of the LS, NS, and MTT tasks were available; thus, it would not have been possible to practice volunteers on this task until performance reached a stable asymptote and still have enough versions of the test available for actual study conduct. In addition, the tests were administered in the same order during baseline and sleep deprived testing which may have resulted in a fatigue effect after 5 hours of continuous testing.

Results of this study raised the question of why (with the exception of decision making performance) SD did not appear to degrade executive function. Results also raised questions regarding both the reliability and validity of the WOMBAT. Further, the validity of the SA construct as a whole has been brought into question. In this dissertation, two approaches were taken to investigate the structure of SA: (1) the relationship between executive function and SA was investigated; (2) change in SA composition following sleep deprivation was examined.

4.10 Relationship Among Executive Functions and SA (Regression Analysis)

The first step toward understanding the SA construct involved investigating the nature of the relationships among executive function and complex task performance requiring SA. To investigate the SA construct, data were analyzed based on the assumption that WOMBAT required SA and that the Overall Score was a measure of ability to maintain SA. In this dissertation it was hypothesized that complex task performance requiring SA would be significantly predicted by multiple measures of executive function because predictive validity is increased by multiple measures (Damos, 1996). To test this hypothesis, two separate multiple

linear regressions were performed (for data sets collected at baseline and again following sleep deprivation) using all cognitive test dependent variables as predictors of WOMBAT overall score. The analyses were run twice—with and without outliers in the data sets.

The linearity and normality assumptions of the multiple linear regression were analyzed post hoc by examining a scatterplot of the data and the normal probability plots of the residuals. Variables were selected using backwards regression. The appropriateness of the variable selection was verified using best subsets regression (See Appendix G).

4.10.1 Complex Task Performance at Baseline, Outliers Present

Backwards elimination selected LNS (working memory), LS (inductive reasoning), and MTT (planning) as significant predictors of baseline complex task performance. The regression equation was: ***Baseline WOMBAT = - 106 + 6.66 Baseline LNS + 8.19 Baseline LS + 4.97 Baseline MTT***. Table 23 lists results of the ANOVA for the baseline regression model with outliers.

Table 23. ANOVA FOR Baseline Regression (includes outliers)

Source	DF	SS	MS	F	P
Regression	3	49903	16634	10.69	0.000
Residual Error	44	68441	1555		
Total	47	118344			

The baseline prediction equation was significant (p=.000), indicating that performance on working memory, inductive reasoning, and planning tasks significantly predicted baseline WOMBAT performance. However, the model accounted for only 32.8% of the variance in WOMBAT performance. The regression residuals were normally distributed (p>.15) and the linearity assumption did not appear to be violated (see Appendix G).

4.10.2 Complex Task Performance at Baseline, No Outliers

When outliers were excluded from the baseline data, the following baseline regression variables were selected: MTT (planning) and LS (inductive reasoning). The resulting regression equation was: ***Baseline WOMBAT = - 60.3 + 5.73 Baseline MTT + 8.36 Baseline LS.*** Table 24 lists results of the ANOVA for the baseline regression model without outliers.

Table 24. ANOVA for Baseline Regression (no outliers)

Source	DF	SS	MS	F	P
Regression	2	32926	16463	9.40	0.000
Residual Error	43	75344	1752		
Total	45	108270			

The baseline prediction equation was significant when outliers were excluded from the data (p=.000), indicating that performance on inductive reasoning and planning tasks significantly predicted WOMBAT performance. In contrast with the first regression equation (baseline, outliers included), LNS (working memory) was excluded from the regression equation. This model accounted for less of the variance (R^2=27.2%) when compared to the first prediction equation (baseline, outliers included). The regression residuals were normally distributed (p=.102) and the linearity assumption did not appear to be violated (see Appendix G)

4.10.3 Complex Task Performance following Sleep Deprivation, Outliers Present

Backwards elimination selected LS (inductive reasoning), MTT (planning), and IGT (decision making) as significant predictors for complex task performance following sleep deprivation. The regression equation was: ***SD WOMBAT = 20.4 + 6.51 SD LS + 4.85 SD MTT - 1.47 SD IGT***. Table 25 lists results of the ANOVA for the regression equation following sleep deprivation.

Table 25. ANOVA for Regression following Sleep Deprivation (includes outliers)

Source	DF	SS	MS	F	P
Regression	3	54729	18243	11.66	0.000
Residual Error	33	51620	1564		
Total	36	106350			

The resulting regression equation was significant (p=.000), indicating that inductive reasoning, planning, and decision making significantly predicted WOMBAT performance following sleep deprivation. The model accounted for a small amount of the variance in WOMBAT performance (Adj. R^2 = 47.0%). An analysis of assumption violations indicated that the regression residuals were normally distributed (p=.074) and the linearity assumption did not appear to be violated (see Appendix G). The sleep-deprived (with outliers) regression model differed from the baseline (with outliers) regression model by selecting decision-making (IGT) in place of working memory (LNS) as a predictor of WOMBAT performance.

4.10.4 Complex Task Performance Following Sleep Deprivation, No Outliers

When outliers were excluded from the data set collected after sleep deprivation, the following regression variables were selected: MTT (planning) and IGT (decision making). The regression equation was: ***SD WOMBAT = 89.1 + 5.33 SD MTT - 1.60 SD IGT***. Table 26 lists results of the ANOVA on the regression model.

Table 26. ANOVA for Regression following Sleep Deprivation (no outliers)

Source	DF	SS	MS	F	P
Regression	2	33336	16668	9.34	0.001
Residual Error	33	58914	1785		
Total	35	92250			

The prediction equation was significant when outliers were excluded from the data (p=.001), indicating that decision making and planning task performance significantly predicted WOMBAT performance after sleep deprivation. The sleep-deprived (no outliers) model accounted for less variance (Adj. R^2=32.3%) than the sleep-deprived (with outliers) model (Adj. R^2=47.0%). The regression residuals were normally distributed (p>.15) and the linearity assumption did not appear to be violated (see Appendix G). The sleep-deprived (no outliers) regression model differed from the baseline (no outliers) regression model by selecting decision-making (IGT) instead of inductive reasoning (LS) as a predictor of WOMBAT performance. Table 27 summarizes the results and shows the variables selected for each regression analysis and the respective adjusted R^2 value.

Table 27. Regression Summary for Complex Task Performance

	Outliers in Data		**No Outliers in Data**	
	Baseline	**Sleep Deprived**	**Baseline**	**Sleep Deprived**
Variables Selected	LNS (working memory) LS (inductive reasoning) MTT (planning)	LS (inductive reasoning) MTT (planning) IGT (decision making)	MTT (planning) LS (inductive reasoning)	MTT (planning) IGT (decision making)
Adj. R^2	38.2	47.0	27.2	32.3

The results supported the hypothesis that multiple measures of executive function significantly predicted complex task performance. The results also supported the notion that the relationship between executive functions and SA changed following sleep deprivation. When outliers were present in the data, decision making (IGT) replaced working memory (LNS) as a

predictor following sleep deprivation. When outliers were excluded, decision making (IGT) replaced inductive reasoning (LS) as a significant predictor following SD. Thus, decision making was a significant predictor of performance on the WOMBAT following sleep deprivation regardless of whether outliers were included in the data. The amount of variance accounted for by the regression equations improved when outliers were included in the data, suggesting that results would be better interpreted when outliers are included in the data sets.

The results indicated that three executive functions were predictive of baseline WOMBAT performance: working memory, inductive reasoning, and planning. After sleep deprivation, working memory was replaced by decision making as a significant predictor of WOMBAT performance. Interestingly, PVT performance was not predictive of WOMBAT performance either during baseline or following sleep deprivation, despite its demonstrated sensitivity to sleep loss. Therefore, while the PVT may be a useful predictor of vigilance decrements, it des not appear to be useful for predicting situational awareness. Assuming that WOMBAT further reflects some relevant aspect of battlefield performance, the findings suggest that measures other than PVT are required to predict operational performance on the battlefield. For example, a handheld assessment of risk-taking propensity (e.g., a Palm Pilot version of the Iowa Gambling Task) may better predict decision making performance on the battlefield than the PVT.

The relatively small percentage of variance accounted for by the regression equations (along with applicability to WOMBAT performance only) precluded generation of strong conclusions regarding the nature of maintaining complex task performance requiring SA under conditions of sleep deprivation. These results appear contradictory to the paired t-test results in which no degradation in working memory was found (however, a trend for degradation of decision making was found). The results also raise questions regarding the mechanisms by which complex task performance requiring SA could be sustained while decision making is degraded.

While WOMBAT performance was predicted by multiple measures of executive function (as hypothesized), there are limitations to the generalizability of this analysis to real-world complex task performance requiring SA. Backwards regression was used for variable selection to keep as many predictors as possible in the model; however, only five executive

function tasks were used as potential predictors. The resulting baseline and SD regression equations accounted for a small amount of the variance in the model (38.2% and 47.0%, respectively). Additionally, the results of this analysis are specific to WOMBAT performance. While some general inferences can be drawn about information processing demands associated with WOMBAT performance and the contribution of executive functions to maintaining SD performance on WOMBAT, the predictive utility of this regression with respect to real-world performance is limited.

4.11 Relationship Between SA and Executive Functions (Baseline Factor Analysis)

To better understand the nature of the relationship between the executive functions tested in this study and SA, factor analysis was used to determine which executive functions load onto the SA construct during baseline testing. Although use of test scores precluded deriving potentially interesting information regarding sleep deprivation-mediated changes in SA, it also initially provided a more straightforward means of evaluating the relationship between executive functions and SA with other factors held constant. Also, to avoid redundancy, dependent variables derived solely for post hoc analyses were not used. The maximum likelihood method with varimax rotation was used. The criteria for selecting the number of factors to interpret was an eigenvalue of greater than one.

Assumptions for the factor analysis were met as indicated by the Kaiser-Meyer-Olin (KMO) measure of sampling adequacy (KMO=.587) and Bartlett's test of sphericity (p=.000) (See Table 28). A KMO value greater than 0.5 indicates the analysis was satisfactory and a significant test of sphericity indicates that the correlation matrix is not an identity matrix (*How to Perform and Interpret Factor Analysis using SPSS*, 2005). Table 29 lists factor loadings of the rotated factor matrix. Factor groupings are shaded. Two factors were interpreted and accounted for 61.3% of the cumulative variance (See Appendix I).

Table 28. KMO and Bartlett's Test for Baseline Factor Analysis

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.587
Bartlett's Test of Sphericity	Approx. Chi-Square	89.236
	df	21
	Sig.	.000

Table 29. Baseline Rotated Factor Matrix

	Factor	
	1	2
Baseline WOMBAT	.085	.905
Baseline LNS	-.021	.416
Baseline LS	-.116	.544
Baseline NS	.932	-.359
Baseline MTT	.862	-.275
Baseline IGT	.370	.293
Baseline PVT	-.411	-.236

The matrix indicated that NS (deductive reasoning) and MTT (planning) both loaded highly onto the first factor. PVT (reaction time) loaded moderately onto Factor 1. WOMBAT (SA) loaded highest on the second factor. Both LNS (working memory) and LS (inductive reasoning) loaded moderately onto Factor 2. Figure 41 shows the baseline factor loading plot.

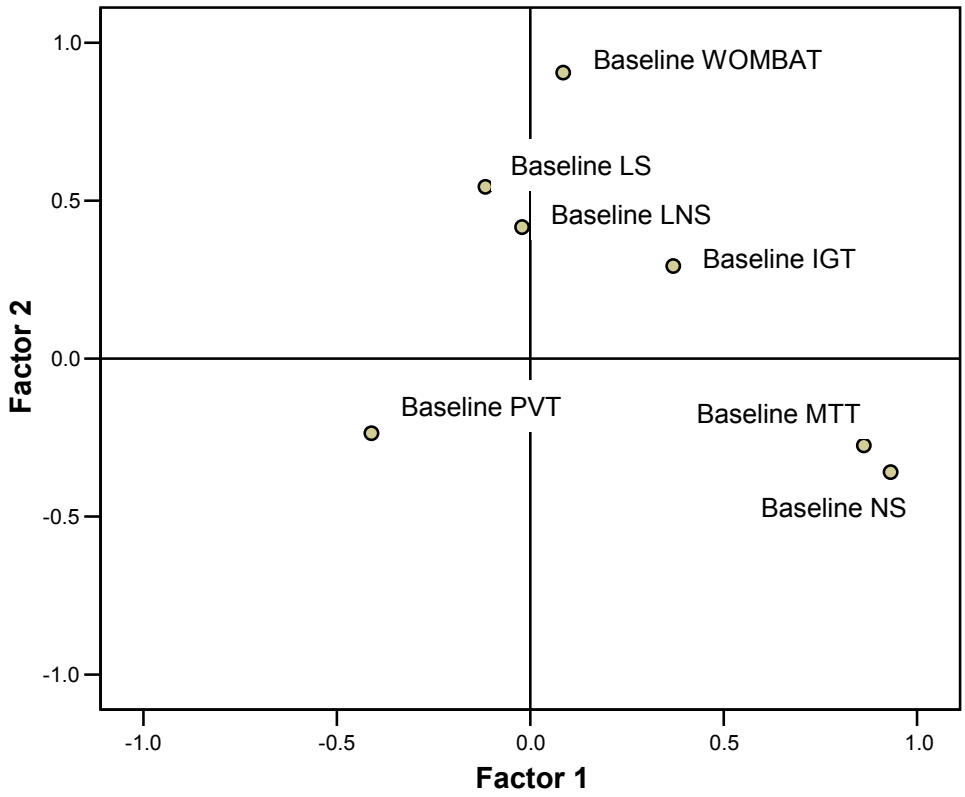


Figure 41. Baseline factor loading plot

The first factor was interpreted as "planning" because NS (deductive reasoning) and MTT (planning) both loaded highly onto the first factor (.932 and .862, respectively). The factor was interpreted as planning rather than deductive reasoning because planning was considered to be a higher level cognitive function than deductive reasoning. PVT reaction time was negatively loaded onto the first factor, indicating that faster reaction times were associated with better planning.

WOMBAT was the only component to load highly on Factor 2 (.905) and therefore, Factor 2 was interpreted as "SA." Results of the factor analysis indicated that two executive functions (working memory and inductive reasoning) also loaded moderately onto SA (.416 and .544, respectively).

4.12 Relationship Between SA and Executive Functions (SD Factor Analysis)

To further investigate the composition of the SA construct, the structure of SA was analyzed under SD conditions to determine whether the executive functions which load highly on the SA construct during baseline changed following sleep deprivation. Factor analyses under baseline and sleep deprivation conditions were compared, and a factor analysis of difference scores was conducted. The maximum likelihood method with varimax rotation was used. The criteria for selecting the number of factors to interpret was an eigenvalue of greater than one.

Assumptions for the factor analysis were met as indicated by the Kaiser-Meyer-Olin (KMO) measure of sampling adequacy (KMO=.649) and Bartlett's test of sphericity (p=.002) (See Table 30). Table 31 shows the factor loadings of the rotated factor matrix. Factor groupings are shaded. Two factors were interpreted and accounted for 54.5% of the variance (See Appendix I).

Table 30. KMO and Bartlett's Test for Sleep Deprived Factor Analysis

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.649
Bartlett's Test of Sphericity	Approx. Chi-Square	44.276
	df	21
	Sig.	.002

Table 31. Sleep Deprived Rotated Factor Matrix

	Factor	
	1	2
SD WOMBAT	.811	-.024
SD LNS	.437	.080
SD LS	.566	.019
SD NS	.248	.419
SD MTT	-.169	.892
SD IGT	.627	.232
SD PVT	-.319	-.287

The matrix indicated that WOMBAT (SA) and IGT (decision making) loaded highly onto Factor 1, followed by moderate loading of LNS (working memory) and LS (inductive reasoning). MTT (planning) loaded highly on the second factor with a moderate loading for NS (deductive reasoning). PVT did not significantly load onto either factor. Figure 42 shows the factor loading plot following sleep deprivation.

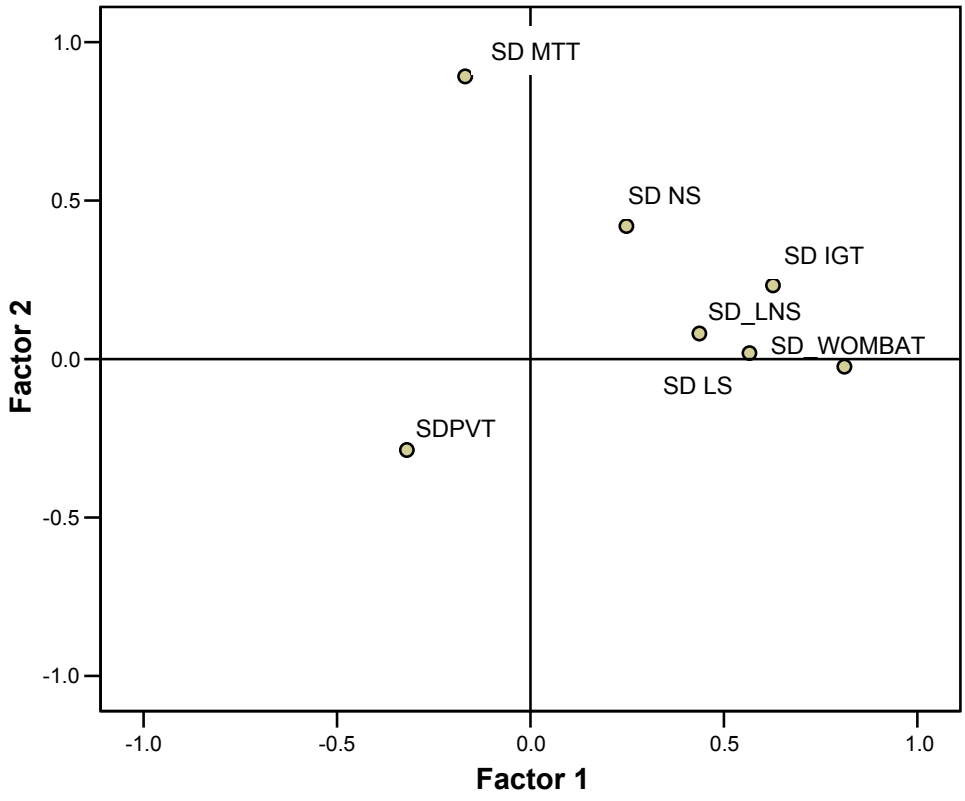


Figure 42. Sleep Deprived Factor Loading Plot

The first factor was interpreted as "SA" because it was the highest loading factor (.811). The second factor was interpreted as planning because it was the highest loading factor (.892). Results of the factor analysis for data collected following sleep deprivation provided initial support for the hypothesis that the executive functions required to maintain SA would change under sleep deprived conditions. The baseline factor analysis grouped LNS (working memory) and LS (inductive reasoning) under the "SA" factor. In comparison, the sleep deprived factor analysis included IGT (decision making) as part of the "SA" factor in addition to LNS and LS. Moreover, total variance accounted for by the "SA" factor increased following sleep deprivation (34.3% after sleep deprivation compared to 27.2% at baseline). Although preliminary, this finding may suggest that additional resources were recruited to maintain SA during sleep deprivation. Finally, in contrast to baseline factor loadings, PVT reaction time (the only lower level cognitive function) was not included in the sleep deprived model because it was the only component not loaded on either factor. This latter finding suggests a disassociation of higher and lower order cognitive functions following sleep deprivation.

4.13 Relationship Between SA and Executive Functions (Factor Analysis of Difference Scores)

To further clarify the relationship among sleep deprivation, SA, and executive function, an additional factor analysis of difference scores (baseline score minus sleep deprivation score) was conducted and included all cognitive tests administered during the study. The goal in looking at the factor structure of difference scores was to see if the cognitive effects of sleep deprivation were differentiated at any level (e.g., higher order effects were separated out from lower level cognition).

The maximum likelihood method with varimax rotation was used. The criteria for selecting the number of factors to interpret was an eigenvalue of greater than one. Assumptions of the factor analysis were violated (KMO=.404; Bartlett's test of sphericity p=.434) (See Table 32). Therefore, drawing strong conclusions from the results were not warranted. Table 28 shows the factor loadings of the rotated factor matrix. Factor groupings are shaded. Three factors were derived and accounted for 57.4% of the variance (See Appendix I).

Table 32. KMO and Bartlett's Test for Factor Analysis of Difference Scores

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.404
Bartlett's Test of Sphericity	Approx. Chi-Square	21.408
	df	21
	Sig.	.434

Table 33. Rotated Factor Matrix of Difference Scores

	Factor		
	1	2	3
IGT Difference Score	.992	-.098	.067
NS Difference Score	.283	.109	-.140
WOMBAT Difference Score	.245	.893	.375
MTT Difference Score	.056	-.466	.159
LS Difference Score	.045	.049	.448
PVT Difference Score	.116	-.044	-.334
LNS Difference Score	.028	.101	-.275

IGT (decision making) was the only component to load highly on the first factor. WOMBAT (SA) loaded highly onto the second factor with moderate loading from MTT (planning). The third factor had only moderate loading from LS (inductive reasoning). Therefore, because no factor loaded highly on Factor 3, Factor 3 was not interpreted. Figure 43 shows the factor loading plot.

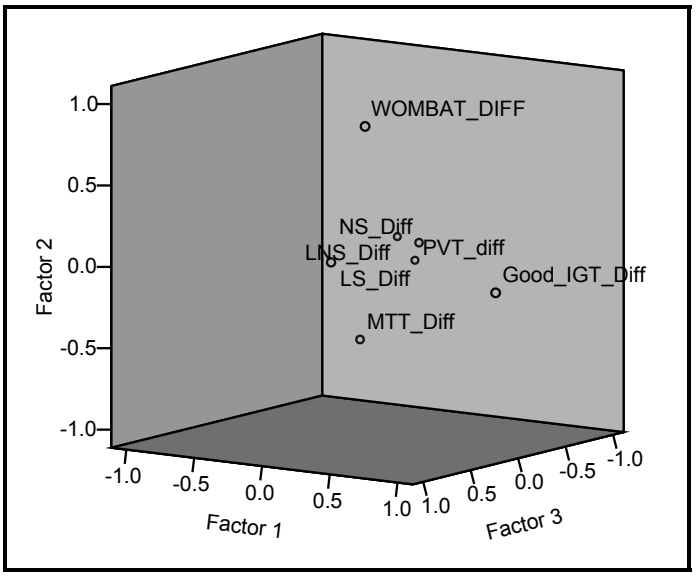


Figure 43. Factor Loading Plot of Difference Scores

Factor 1 was interpreted as "decision making" because IGT (decision making) was the only test to load onto the first factor (.992). Factor 2 was interpreted as "SA" because WOMBAT (SA) loaded highest onto the second factor (.893). Interestingly, the factor analysis of difference scores separated decision making and complex task performance (SA/planning/inductive reasoning) into two factors. This finding suggests that performance degradation was not due to degradation of information processing as a whole but rather due to degradation in the decision making and SA components.

In Chapter 2, a process approach was proposed for defining and ultimately measuring SA. Several hypotheses about the SA construct were proposed as follows: (1) SA is not a construct separable from executive function, decision making, and action; and therefore, (2) no single measure of cognition would significantly predict complex task performance requiring SA. The regression and factor analysis results of this study potentially support -- and refute -- these hypotheses, as discussed in the following sections.

4.14 Discussion of SA as a dynamic process

The hypothesis that SA is a dynamic process was supported by a comparison of baseline and sleep deprived factor analysis results, in which it was shown that factors predicting performance changed from baseline to sleep-deprived conditions. The most widely cited

definition of SA is that of Endsley (1995, p. 36), "…the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." (Endsley, 1995a) argued that SA is a state of knowledge separate from decision making and action and that outcome performance is only an indirect measure of SA and should be considered separately. Despite numerous references to Endsley's definition of SA, neither this definition nor any other operational definition of SA has been validated. Conceptions of SA range from domain-specific definitions that pertain to only one system to broad, abstract concepts that encompass any cognitive activities necessary for performance in a dynamic environment (Durso and Gronlund, 1999). When the various definitions of SA are considered as a whole, SA can be described as "all knowledge that is accessible and can be integrated into a coherent picture, when required, to assess and cope with a situation" (Sarter and Woods, 1991, p.55). Rather than debate the merits of one definition of SA compared to another, the goal of this dissertation was to take a first step toward understanding how SA can be understood and measured in a way that will ultimately increase the ability to predict operational performance.

In a dynamic environment, the cognitive processes (including SA) required to sustain performance adapt to task demands such that the relationship among SA, decision making and action is continuously shifting (Durso and Gronlund, 1999). A comparison of baseline versus sleep deprived factor analysis results indicated that the structure of complex task performance changed in response to SD. The two factors that emerged from both analyses were interpreted as "planning" and "SA". However, not only did the composition of the planning and SA factors change after SD, but the relative loadings and the amount of variance contributed by the factors changed as well. The baseline SA factor had moderate loading from working memory and inductive reasoning. After SD, decision making was also loaded on the SA factor. In the baseline matrix, planning accounted for 34% of the variance compared to 20% after SD. Similarly, SA accounted for 27% at baseline compared to 34% after SD. Thus, the relative contribution of planning decreased in response to SD and the relative contribution of SA increased after SD. When taken together, the relationship among information processing components changed in response to task and environmental demands. The results supported the proposal that SA is not a permanent state when considered in the context of dynamic complex task performance (Taylor, 1990). Therefore, the *processes* involved in adapting to dynamic SA

and task requirements must be understood in order to predict real-world performance. In sum, if the nature of complex task performance is dynamic, the conception of SA must be process-oriented.

Additionally, the results of baseline and sleep deprived factor analyses provided some initial support for the hypothesis that SA was not a construct entirely separable from information processing mechanisms. Although results from the baseline factor analysis indicated a separation of planning/deductive reasoning from situation awareness, results also indicated that working memory and inductive reasoning loaded moderately on to the SA factor. The additional loading of decision making to the SA factor in the sleep deprived analysis provided further support for the link between SA and other information processing components. And finally, the factor analysis of difference scores included planning and inductive reasoning as a component of SA. When these results are considered jointly, SA was not entirely separable from information in any analysis despite the varied composition of SA in each of the separate analyses. Thus, these results support the hypothesis that SA is not separable from information processing as a whole and further support the assertion that making a clear distinction between the state of awareness and the processes used to achieve and update the state of awareness has not been shown to be scientifically defensible (Adams, Tenney, and Pew, 1995).

4.15 SA, Decision Making, and Performance Prediction

Taking the previous hypothesis a step further, an extension to (Wickens and Hollands, 2000) SA/Information Processing model was proposed to include decision making as a component of the SA process. Considering SA outside the context of information processing may be necessary to determine where there may be disconnects among (a) the presentation of information; (b) processing of information; and (c) resulting actions. However, a direct causal relationship does not necessarily exist between SA and objective performance (Flach, 1995). Evidence has shown that operators use SA differentially based on experience (Randel and Pugh, 1996). Additionally, good decisions and good performance can be achieved with poor SA while poor decisions and poor performance can be achieved with good SA (Wickens, 2002a). Thus, the measurement of SA by itself is not sufficient for predicting performance.

Results of the sleep deprived factor analysis indicated a separation of SA and decision making under sleep deprived conditions and thus, did not support the hypothesized extension of

Wickens and Hollands (2000) model to include decision making as part of the process of maintaining SA. However, this finding does not refute the hypothesis that SA is not separable from information processing as a whole. It does provide some initial evidence that SA and decision making are not the same construct. Results from the difference score analysis indicated that decision making was a separate factor from the other executive functions. Thus, while decision making and SA are distinguishable, decision making still appears to contribute to producing and predicting performance outcomes. Baseline and sleep deprived regression results showed that planning and decision making significantly predicted complex task performance requiring SA. The results of the regression only serve to provide further support that complex task performance cannot be predicted by SA alone – that is, in this instance SA did not degrade – thus, the assumption based on SA as a predictor would have been that performance also did not degrade. In fact, there was a trend towards degraded performance outcomes as a result of degraded decision making rather than degraded SA. Because the two functions cannot be separated in a complex task performance context, they should not be separated in the prediction of complex task performance.

In summary, the results of this study served as initial steps to better conceptualize the SA construct so that complex task performance requiring SA could be better understood and eventually better predicted. The results did not directly support or refute an operational definition of SA as a state of knowledge; however, the results did suggest that SA is a dynamic process that adapts to changing task demands and environmental stressors. The results of this study further indicated that performance on a task requiring SA was not separable from other information processing components including working memory, reasoning, and decision making regardless of the environmental conditions. And finally, despite the separation in constructs between SA and decision making from a prediction standpoint, results showed that complex task performance cannot be predicted by SA alone.

4.16 Limitations of SA analysis

The interpretation of factor analysis results is subjective and the factor analysis of the SA construct was limited in several ways. First, relatively few dependent variables were used in the model. It is therefore likely that latent variables (i.e., variables not measured) contributed to WOMBAT performance. These include (but are not limited to) visual perception, risk-taking,

and motivation. To account for these and other latent variables in a model of SA, more tasks measuring these variables should be included. Second, WOMBAT was hypothesized to measure SA; however, WOMBAT has not been sufficiently validated as a measure of SA. Although the data provided initial support for a process definition of SA, the results and conclusions drawn using data from the performance-based WOMBAT methodology may not be conclusive enough to necessarily distinguish SA from executive function or complex task performance. And finally, because the decision making task (IGT) was given 3.5 hours prior to administering WOMBAT and the other cognitive tasks, the results of this study may not warrant forming strong conclusions about the relationship between decision making and other information processing components.

4.17 Study Contributions to SA research

Although SA has received much attention as a causal factor in aviation accidents and as a predictor of performance in complex systems, much of the evidence is anecdotal and no consensus has been reached in the scientific community on how to conceptualize or define SA. Thus, this dissertation took an exploratory approach to studying the SA construct by investigating the relationship between information processing and SA as well as analyzing an existing SA measurement technique (i.e., WOMBAT). This dissertation contributed to the body of SA literature by identifying gaps between the numerous existing SA definitions and models and attempting to bridge that gap by testing the hypothesis that SA is a dynamic process.

Results of this study provided initial support for a dynamic, process definition of SA and also concluded that SA should be studied within the context of information processing as a whole rather than as a separable construct. Furthermore, the results of this study provided data to support anecdotal evidence that SA is not the only factor affecting operational performance (Marsh, 2000). Thus, the major contribution of this study was to clarify and emphasize the need for further validation of the SA construct at a more fundamental level as well as to suggest that future research of SA use a process-oriented approach. In other words, the results of this study suggested that the parameters of SA need to be clarified and their relationship to one another in the context of complex task performance need to be better understood in order to improve SA measurement and performance prediction. Furthermore, this dissertation was able to make suggestions for developing an "ideal" SA measurement technique and engineering

recommendations for improving existing SA measurement techniques. The next section presents several proposed characteristics of an ideal SA measurement system based on the advantages and disadvantages of existing SA measurement methodologies. An example is then provided to illustrate how these characteristics can be applied to improving the performance-based WOMBAT SA measurement system.

4.18 Engineering Recommendations

4.18.1 SA Measurement Recommendations

Recommendations for improving SA measurement were based on the advantages and disadvantages of existing SA measurement techniques as well as the results of this study. It is critical to note at this point that this dissertation is not arguing that WOMBAT did not require SA for good performance, but rather is suggesting that the measures embedded in WOMBAT did not measure SA directly. Therefore, this discussion equated performance on WOMBAT with performance on a task requiring SA.

There are three main categories of SA measurement techniques in use today: query, rating, and performance-based. Query techniques involve asking the operator directly about his or her perception of a situation; Rating techniques ask either observers or operators to rate their situation awareness; and Performance-based SA techniques infer the level of SA from performance outcome measures. Performance-based techniques are based on the premise that SA is necessary for good performance and that the level of SA achieved can be inferred from performance (*The development of situation awareness measures in ATM systems*, 2003). Each measurement technique provides advantages and disadvantages for measuring SA and the inadequacies in one method are often compensated for by the advantages of another method. Ideally, an SA measurement technique would incorporate the advantages of each SA measurement approach (query, rating, and performance-based) while moderating the disadvantages. Thus, an ideal measurement technique would include knowledge measures, process measures, and outcome measures. The recommendations in this section attempt to bridge the gap between the three approaches to SA measurement and mitigate the limitations of each. Table 34 shows the relative advantages and disadvantages of each approach.

Table 34. Advantages and Disadvantages of SA Measurement Approaches

	Subjective Rating	Query	Performance-Based
Advantages	• Easy to use • Useful for simulations and real-world tasks • Do not require complex customization • Not intrusive	• Quantitative results • diagnostic	• Not intrusive • Consider contribution of other elements to performance
Disadvantages	• Cannot measure subconscious knowledge or behavior • Can be influenced by mission outcomes or operator performance • Potential confound with workload	• Cannot measure subconscious knowledge or behavior • Interrupt task flow • Require customization for specific domains • Limited predictive validity	• Cannot separate SA from other contributors to performance •

Subjective rating techniques are easy to use, useful in a wide range of both real-world tasks and simulations, do not require customization for use in different domains, and are not intrusive (The development of situation awareness measures in ATM systems, 2003; Endsley et al., 1998). Query techniques yield quantitative, objective results and can isolate system design elements that may decrease operator SA (Endsley et al., 1988; Endsley, 1988). Performance-based techniques are not intrusive and consider the interaction of multiple cognitive processes involved in performance requiring SA.

Despite the advantages of each approach, there are several limitations. The process of maintaining SA and resulting performance is not always a conscious process and experienced operators will often process information that is not directly observable or part of conscious attention allocation (Hartman and Secrist, 1991). Thus, operators may not be aware of their own SA despite good performance. This potential confound significantly limits the predictive validity of rating and query techniques. The operator may know he is lacking knowledge of the situation or environment, but not know the extent of his reduced SA. Or because of his lack of SA, he may think he has perfect SA when that is not actually the case (Endsley, 1988). Thus, if the operator is unaware of his own subconscious knowledge, neither subjective ratings nor direct queries will capture that information. Although performance-based techniques do not require the

operator's conscious awareness of his or her own SA, good performance can result from poor SA and poor performance can result with good SA (Endsley, 2000). Therefore, performance-based techniques cannot always distinguish the root causes of performance problems and may not be specific enough to determine whether SA or other information processing components were affecting performance (e.g., WOMBAT overall score). When subjective self-ratings occur at the end of a scenario, the operator's perception of SA may be influence by the outcome of the mission. For example, if the mission was successful, the operator might rate himself as having high levels of SA. Conversely, if the mission outcome was unsuccessful, the operator might rate himself has having low SA. Another disadvantage of subjective rating techniques is the potential confounding of SA with subjective workload (Endsley, 1995). And finally, query techniques interrupt the natural flow of the task and require an extensive requirements analysis to provide comprehensive SA information (*The development of situation awareness measures in ATM systems*, 2003). Furthermore, discrete measurements of SA (i.e., measures of SA as a state of knowledge at a given point in time) do not take the dynamic nature of complex tasks into consideration and thus, may not be sensitive to or predictive of real-world performance.

Based on the advantages and disadvantages of each approach described above, this dissertation proposed two general recommendations for moving toward a better SA measurement technique. To account for the dynamic relationship between SA and other information processing components, SA measures can be derived from an operationally relevant process definition of SA. To improve the diagnostic capability of a performance-based approach, knowledge measures and subjective assessments of SA can be included in addition to performance outcome measures so that SA can be better distinguished from other cognitive processes contributing to performance. Because measures that are task specific may provide incomplete information about SA (Uhlarik and Comerford, 2002), it would be advantageous to include both low-level and high-level performance outcome measures. An example of a low-level metric is reaction time to identify a target on the grid. An example of a high-level metric is total number of friendly targets identified. It is important to note that embedded SA metrics should be process-oriented. Furthermore, metrics should be embedded within the system to avoid interrupting task performance and thus, disrupting the process of maintaining SA with artificial tasks while attempting to measure it. The prediction of performance on tasks requiring SA should be robust to dynamic conditions, but sensitive to small changes in performance at the

same time. Therefore, measures should be repeated at frequent intervals. If possible, performance measures should adapt to the user's actions. And finally, metrics should then be included within a larger set of task-dependent performance metrics (e.g., action selection, decision making) to predict performance outcomes.

A first step toward incorporating the above recommendations would be to develop an operationally relevant process definition of SA. From that definition, a general SA taxonomy could be created and both general and task-dependent SA process metrics could be derived for a specific system under investigation. Using the SA definition proposed in Chapter 2, the SA measurement taxonomy would have seven main SA categories from which metrics could be derived. They are as follows:

- Perceiving Information
- Shifting and Sustaining Attention
- Allocating Priorities
- Reasoning
- Planning
- Decision making
- Coping with Stressors

To analyze SA while operating a system similar to WOMBAT, several task-specific SA process metrics could be derived. For example, under the "Perceiving Information" category, a metric called "Awareness of Missing Target" could be derived and measured as follows: time elapsed between onset of target disappearance and activation of the grid to search for missing target. This metric is not a direct measure of the state of knowledge that a target has gone missing, but rather a measure of how well SA was maintained during a complex task, how quickly the operator was able to switch attentional resources, and whether the appropriate action was selected. This metric provides more valuable information than simply measuring the knowledge that the target has gone missing because it also provides an indication of the operator's ability to respond quickly and appropriately to that information. Importantly, the appropriateness of the response (i.e., short vs. long reaction time) would be task-specific. For example, a long reaction time to critical information might indicate poor SA, but a long reaction time to non-critical information might indicate good awareness that the information does not require immediate attention. Therefore, caution should be taken when evaluating and interpreting performance

metrics. A one-size-fits-all approach to defining and measuring SA performance may result in misleading interpretations of data.

To improve the diagnostic capability of this particular measure and determine whether SA or decision making played a larger role in performance, a knowledge metric could easily be embedded in the program to separate knowledge from decision making and action. For example, requiring the operator to acknowledge the target as critical or non-critical by clicking a button would provide reaction time data as well as SA data. Subjective ratings of SA can also be included at the end of the mission or simulation to further improve the diagnostic capability of the SA measurement technique. Although this is a simplistic example of integrating process, knowledge, and subjective metrics into one technique, this methodology could be extrapolated to larger, more complex systems as either an addition to existing techniques or as an alternative approach.

There are limitations to this process approach as well. First, it would be impossible to identify every potential SA subtask an operator may encounter for a large-scale, complex system in which task demands and SA requirements change rapidly. Thus, it would be important to develop a breadth of higher level SA metrics in addition to lower level metrics. Second, it may be difficult to embed measures into existing software systems already in operation on the battlefield. However, embedded measures can potentially be integrated into new large scale systems using emerging technologies such as decision aiding and dynamic function allocation which already handle real-time updates of large amounts of information and adapt to user's actions. Third, some researchers might argue that performance outcomes on SA subtasks (i.e., SA process metrics) are an indirect measure of SA and may be confounded. This "ideal" approach attempts to refute that argument based on the proposed process definition of SA. The proposed recommendations for improving SA measurement were developed by incorporating the potential confounds into the metrics at a level which better predicts outcome performance rather than attempting to eliminate all confounds and thereby reducing predictive validity within a dynamic, complex task.

4.18.2 Improving Existing SA Measurement (An Exemplar using WOMBAT)

This section provides a specific example for applying the recommendations presented in the previous section to an existing SA measurement technique (WOMBAT). The suggested modifications to WOMBAT were based on validity issues noted during study procedures, results

of data analyses, and theoretical advantages and disadvantages of performance-based SA measurement techniques.

4.18.2.1 WOMBAT Validity as a Measure of SA

WOMBAT has several advantages for use as an SA measurement technique. WOMBAT has face validity as a task that requires SA for good performance. Although relatively simplistic in design, the primary tracking task simulates air traffic control requirements for monitoring target paths, predicting collisions between targets, and finding "missing" targets. In addition, WOMBAT requires the operator to perform secondary tasks (e.g., figure rotation, two-back digit canceling, and quadrant location) while simultaneously monitoring the status of the primary tracking task. The system uses a process-oriented approach to defining SA and a performance-based approach to measuring SA. Moreover, WOMBAT is not domain-specific and thus, does not require trained operators. And finally, studies have shown that WOMBAT scores are not correlated with computer or computer game experience (Roscoe, Corl, and LaRoche, 2001). Despite its advantages, relatively few studies have been conducted to support content and construct validity.

A factor analysis of all meaningful WOMBAT dependent variables was conducted to determine what factors were present in the WOMBAT scoring algorithm. Results indicated that WOMBAT may not measure SA, but rather it may measure complex task performance. Principal components method was used with varimax rotation because factor analysis would not converge using maximum likelihood method. The assumptions for the principal components analysis were met (KMO=.617; Bartlett's test of sphericity p=.000) (See Table 35). Table 36 shows the factor loadings. Factor groupings are shaded.

Table 35. KMO and Bartlett's Test for Factor Analysis of WOMBAT Dependent Variables

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.617
Bartlett's Test of Sphericity	Approx. Chi-Square	1319.031
	df	78
	Sig.	.000

The factor analysis yielded three factors which accounted for a cumulative 80.8% of the

variance in the model. The first factor was composed of all dependent variables relating to bonus tasks and was interpreted as "bonus tasks performance." All the dependent variables that loaded highly onto the second factor were measures of tracking performance. Therefore, the second factor was interpreted as "tracking performance." The only dependent measure to load onto the third factor was a measure of visual search time and was interpreted as "visual search." Therefore, the separation of WOMBAT measures into three distinct factors (bonus, tracking, and visual search) provided a strong indication that the WOMBAT scoring algorithm was not measuring some overarching SA construct, but rather was measuring complex task performance.

Table 36. Rotated Factor Matrix of WOMBAT Variables

	Component		
	1	2	3
Total Tracking	.083	.916	-.288
Overall Tracking Percent	.078	.870	.115
Total Collision Detection	-.078	.738	-.396
Total Solid (Figure Rotation)	.891	.114	-.105
Total Quadrant	.884	.145	.170
Total Sequences Mastered	.744	.262	.038
Total Two Back	.832	.052	.211
Total Bonus	.983	.118	.093
Total Overall	.499	.819	-.236
Number of Missing Targets Found	.095	.882	.264
Time Before Missing Target Found	.177	-.184	.888
Visible Collisions	.250	.660	-.301
Correct Figure Rotation Answers	.905	-.011	-.072

This dissertation is not asserting that SA was not required for good performance on WOMBAT, only that the scoring algorithm did not measure it directly. For example, a potentially useful measure of SA would have been a measurement of the time it took for the participant to notice that a target had gone missing. A dependent variable with this name "time before missing target found" was collected by the system, but it was a measurement of the time it took for the participant to find the missing target once the participant activated the grid, not a measurement of the time from when the target initially went missing until the participant activated the grid to begin searching for it. Therefore, the participant's SA was not measured,

but rather his/her visual search time was measured. Unfortunately, the additional dependent variables automatically collected by WOMBAT did not appear to be more useful measures of SA than overall score.

The validity of some of the subtasks is in question as well. For the computerized working memory task (two-back digit canceling), participants often simply held their fingers on the correct keys until it was time to press them rather than storing the digit in their working memory. Additionally, the order of problems presented in the quadrant location task and the figure rotation task were not randomized within the testing session, so order effects during the SD session cannot be ruled out. The sensitivity of WOMBAT to small changes in performance was discussed earlier, but applies to the validity of the test as well. An inspection of the mean interval scores for baseline and sleep deprived testing sessions indicates that the pattern of the scoring profile is very similar – and both times was characterized by an increase from intervals 1 to 3 followed by a scoring decline between intervals 6 and 7 and then another scoring increase through interval 12. It is possible that this pattern reflected a "forcing function" in the scoring algorithm. In other words, the score may have been more a function of a delayed software response to the user's actions than of actual changes in the user's performance. Because it was determined that WOMBAT overall score was not a good measure of SA, a structural equation model of SA was not appropriate. More investigation of the construct of SA is needed before a model can be built.

4.18.2.2 Engineering Recommendations for Improving WOMBAT

Specific recommendations for improving WOMBAT fall under two categories: (1) SA measurement improvements and (2) software improvements. The premise behind WOMBAT is that the operator must remain aware of the state of the primary tracking task while performing secondary tasks and then respond appropriately to changes in the situation. The software program has face validity for requiring situation awareness and simple modifications could improve both its measurement of SA and its usability. First, meaningful SA process metrics could be integrated into the existing software by adding time stamps for particular user actions (e.g., time elapsed between target's disappearance and operator's activation of grid to search for missing target). Secondly, a questionnaire can be administered at the end of the instruction period, but before the test session begins to assess the operator's understanding of the "rules" for achieving good performance. Thus, the operator's strategy for taking the test can be analyzed as

a potential confound for either good or poor performance. Third, operators could complete a brief computerized questionnaire at the end of the testing session to subjectively rate their SA during the task.

Several WOMBAT software issues also need to be addressed. The software should be recoded to improve response time to participant actions. For example, participants were instructed to monitor task worth indicators continuously (to maintain SA) and then make appropriate task selections based on the relative task worths. However, the worth indicators frequently disappeared from the screen and often did not change values in real-time to reflect an event had occurred. As an example, when a target went missing, the bonus task worth indicator should have shown a decrease in bonus task worth and the tracking task indicator should have shown an increase in tracking task worth (as per the WOMBAT manual). The participant should have maintained awareness of the situation, noticed the relative change in task worths, and then responded by trying to find the missing target. More often than not, however, the worth indicators were not updated real-time. Moreover, despite the software delay in updating the worth indicators, the scoring algorithm calculated the participant's response time (i.e., SA) using the time stamp from the event rather than the time stamp from when the worth indicator was updated. As a result, the participant's SA score was penalized for not responding to the event when it was possible that the participant did respond quickly (and maintained SA) in response to the worth indicator (as instructed) when it was eventually updated to reflect the event. Thus, not only was the participant's score inaccurately measured, but the scoring algorithm was confounded by a conflict between explicit instructions given to the participant and the software's delayed updates. It is critical to validate WOMBAT's scoring algorithm from both a programming and SA measurement perspective to ensure that the performance scores accurately reflect the operator's ability to maintain SA.

4.19 Future SA Research

Although this dissertation made some preliminary recommendations for developing an "ideal" SA measurement technique and made specific suggestions for improving WOMBAT, additional SA research is needed to further validate the SA construct. Current SA research focuses primarily on developing SA measurement methodologies from existing definitions and conceptions of SA. As a result, there are numerous methodologies based on numerous

conceptualizations that have, to date, resulted in little predictive validity. If the ultimate goal of defining and measuring SA is to improve predictive validity for operational performance, perhaps SA research needs to refocus its efforts on addressing the basic research questions that so far have remained unanswered or under debate. These include, but are not limited to the following:

- What are the parameters that define SA?
- Are some parameters more important than others?
- Can a hierarchy of parameters be created?
- Are the parameters consistent across different domains?
- What parameters are context-specific?
- What is the relationship between these parameters in a dynamic, complex task environment?

To address these basic questions and to move toward improving the predictive validity of SA measurement, this dissertation proposed a future research agenda. The first recommended step is to conduct basic research to create and validate a hierarchy of SA parameters from which tailored tools can be developed. Research should then validate these tools for both construct and predictive validity in simulations and operational environments. Next, the aggregate research findings can be consolidated to determine which parameters or elements of SA may consistent across multiple domains and which may be context-specific. And finally, the research can move toward developing tools that are valid in both simulations and operational environments and significantly predict operational performance.

This research agenda need not occur separately from ongoing SA research. Rather, it can be conducted concurrently so that research findings can be consolidated and analyzed to form a clearer vision of the "big picture." Thus, the results of this study and the engineering recommendations presented in this dissertation can be applied to both basic and applied SA research in the pursuit of a better understanding of SA and better predictive validity.

4.20 Future Sleep Deprivation Research

This dissertation contributed to the larger body of sleep deprivation literature by attempting to confirm and refute several hypotheses about the effects of sleep deprivation on executive function. The results of the study confirmed the degradation of sustained attention as a

result of sleep deprivation, but also indicated that there is inconsistency in research findings across multiple sleep deprivation studies on executive function. Although the results of numerous studies have consistently provided evidence that sleep deprivation degrades attention, the effects of sleep deprivation on executive function and complex task performance are not well understood. Thus, designing and testing engineering solutions to mitigate the effects on these types of tasks will be more difficult. For example, the results of this study showed a trend toward degraded decision making; however, the effect was a function of time on task. In operational settings, the effects of sleep deprivation may also be time and task-dependent. Therefore, there will not likely be a universal solution to mitigate the effects of sleep deprivation on executive function. As research on neurophysiological effects and individual differences continues, research designed to develop and test engineering solutions should progress jointly. The results of engineering studies may provide insight into behavioral responses to sleep deprivation as well as insight into how best to account for the degrading effects of sleep deprivation on performance. Thus, the results of this dissertation reemphasized the need to identify potential neurophysiological correlates to behavioral performance and to clarify the factors affecting individual differences in resiliency to sleep deprivation.

The results of numerous brain imaging studies (e.g., Drummond et al., 2000; Chee and Choo, 2004; Drummond and Brown, 2001; Gosselin, Koninck, and Campbell, 2005; Killgore, Balkin, and Wesensten, in submission) have indicated that relative activation of brain regions may account for or at least partially explain the cognitive performance response to sleep deprivation. However, more brain imaging studies are needed to investigate the neurophysiological and behavioral correlates of performance under sleep deprived conditions such that the mechanisms are better understood, performance prediction can be improved, and ultimately, design solutions can be developed to mitigate the effects.

In addition, the inter-individual differences that appear to determine a person's resiliency to sleep deprivation should be further studied. To date, results of sleep deprivation studies have failed to find consistent or definitive demographic characteristics of individuals that account for inter-individual differences in response to sleep deprivation, nor have results shown that an individual's sleep history can account for these differences (Akerstedt, 1999; Kerkhof, 1985). Thus, future research on individual differences in response to sleep deprivation should focus on

determining whether there is a neurobiological basis for inter-individual differences (Van Dongen et al., 2004).

As a corollary to continued research on the cognitive effects of sleep deprivation, potential engineering solutions for mitigating the effects of sleep deprivation on cognition and performance should be tested for effectiveness. Some initial research has already been conducted on improving vigilance under sleep deprived conditions. Results of an Army study on the effectiveness of auditory vs. visual helmet mounted displays (HMD) in facilitating navigation for sleep-deprived soldiers indicated that performance improved using either modality, but that the auditory modality was less affected by sleep deprivation as time on task increased (Brown, 2004). This finding was consistent with a study of sleep-deprived driving performance which found that driving performance was not degraded by sleep deprivation when operators interacted with a vehicle information system, regardless of whether the information was presented auditorily or visually (Lee, Dingus, Mollenhauer, Brown, and Neale, 1997). Additional studies have found that operators were better able to sustain attention when information was presented in the auditory modality (as cited Lee et al., 1997)). Thus, to counteract the effects of degraded vigilance due to SD, system designers should research the effects of multiple modalities on sustained attention with respect a specific system.

Studying sleep-deprived performance on complex tasks is a complicated undertaking because the effects of sleep deprivation on complex task performance occur at many levels. The effects of sleep deprivation are a function of time on task, level of task (primary vs. secondary), and task workload. Although there was no time-on-task effect for WOMBAT in this study, Lee et al. (1997) driving study results indicated that as time on the driving task increased, driving performance degraded. Interestingly, task difficulty was a factor in degraded performance only in the last third of the simulation (i.e., after time on task exceeded 60 minutes). WOMBAT results did not indicated a degradation in primary or secondary tasks; however, results of a Brown (2004) helmet-mounted display study found that secondary task performance (detection of enemy soldiers) degraded as a result of sleep deprivation even though the primary navigation task performance improved. Hockey, Wastell, and Sauer (1998) investigated the effects of sleep deprivation on performance of a complex, multilevel compensatory control task using both machine-centered interfaces and human-centered interfaces. Results indicated no degradation in primary task performance (maintaining three key system variables within target ranges) on either

interface. However, on the machine-centered interface, there was significantly degraded performance on secondary tasks (acknowledging alarms, recording tank level, and logging system failures) as well as a degraded ability to change decision making strategies. Thus, the effects of sleep deprivation can be seen on secondary task performance even if it appears that primary task performance is unaffected. Furthermore, results of additional studies have shown that increasing task workload mitigated the effects of sleep deprivation on performance until task overload occurred (as cited in Chee and Choo, 2004). Thus, some research indicates that increasing workload to a certain threshold can reduce the effects of sleep deprivation on performance. In sum, results from sleep deprivation studies investigating complex task performance have shown mixed results. Therefore, before the effects of sleep deprivation on complex task performance are studied further, it is critical to understand the mechanisms by which complex task performance requiring SA is maintained under baseline conditions.

4.20.1 Methodological Recommendations for Future Sleep Deprivation Research

Methodological issues with task selection, data analysis, participant selection, and study procedures should be taken into consideration for future sleep deprivation studies investigating executive function tasks. In general, a real-world simulation task should be included in addition to traditional psychological tests to investigate the generalizability of executive function results to simulated operational settings. However, it is important to validate and/or pretest the simulation prior to the study to ensure the dependent variables are meaningful and sensitive to sleep deprivation. In addition, alternative statistical analyses should be considered to account for trends or effects not found in paired t-test or ANOVA analyses (e.g., post hoc analysis of IGT data). Furthermore, asking participants about task performance strategies may provide insight into the effects of sleep deprivation that would not otherwise be apparent based solely on objective dependent variables. In addition, subjective data on participant strategy may identify potential confounds in performance data. For example, the underlying cause of poor performance score may stem from a misunderstanding of task rules rather than an inability to perform the task due to degraded SA. To address potential learning effects on cognitive tasks and to help determine the size of the sleep deprivation effect on cognitive performance, a fully rested control group may be included in the study design. Additionally, if online maintenance of wakefulness tests (MWT) are not feasible, removing the MWT from the study procedures may be warranted to prevent participants from getting small amounts of sleep at regular intervals

throughout the sleep deprivation portion of the study. And finally, the principal investigator may want to consider excluding participants for whom English is a second language to avoid a potential confound on cognitive tasks with a large verbal or reading component (e.g., deductive reasoning).

4.21 Conclusion

The purpose of this dissertation was to form a better understanding of the effects of sleep deprivation on real-world, complex task performance requiring SA and to simultaneously examine the construct validity of SA. Results of the study indicated no degradation of higher order cognitive function as a result of sleep deprivation, but showed a trend toward degraded decision making and planning. Additionally, the results of this study provided initial support for a dynamic, process definition of SA and also indicated that SA should be considered within the context of information processing as a whole rather than as a separable construct. This dissertation was exploratory in nature and illustrated the complexity involved in understanding and predicting operational performance. Although the findings of this study provided insight into the link between sleep deprivation, executive function, SA, and complex task performance, more questions were ultimately raised than answered. The potential lack of internal validity in this study prevented drawing strong conclusions about the nature of the SA construct; therefore, future research directions were proposed to further explore both the cognitive effects of sleep deprivation and the construct of SA. This dissertation was a small step toward merging and understanding the relationship between two separate, but operationally relevant issues (sleep deprivation and situation awareness) with the goal of ultimately improving the prediction of operational performance requiring SA under environmental stressors.

References

Adams, M., Tenney, Y., and Pew, R. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85-104.

Akerstedt, T. (1991). Sleepiness at work: effects of irregular work hours. In T. Monk (Ed.), *Sleep, Sleepiness, and Performance* (pp. 129-154). Chichester: John Wiley and Sons.

Akerstedt, T. (1999, June 3-4). *Individual Differences in Reaction to Irregular Work Hours.* Paper presented at the RTO HFM Workshop on "Individual Differences in the Adaptability to Irregular Rest-Work Rhythms/Status of the Use of Drugs in Sleep-Wakefulness Management", Venice, Italy.

Albers, M. (1999, October). *Information design considerations for improving situation awareness in complex problem-solving.* Paper presented at the 17th annual international conference on Computer documentation, New Orleans, LA.

Axelrod, B. (2002). Validity of the Wechsler abbreviated scale of intelligence and other very short forms of estimating intellectual functioning. *Assessment, 9*(1), 17-23.

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., et al. (2000). *Effects of Sleep Schedules on Commercial Motor Vehicle Driver Performance* (Technical Report No. DOT-MC-00-133). Silver Spring: Walter Reed Army Institute of Research Division of Neuropsychiatry.

Bandaret, L., Stokes, J., Francesconi, R., Kowal, D., and Naitoh, P. (1981). Artillery teams in simulated sustained combat: Performance and other measures. In L. Johnson, D. Tepas, W. Colquhoun and M. Colligan (Eds.), *Biological Rhythms, Sleep, and Shiftwork* (pp. 459-477). New York: Spectrum.

Barfield, W., Rosenberg, C., and Furness, T. A. I. (1995). Situation Awareness as a Function of Frame of Reference, Computer Graphics Eyepoint Elevation, and Geometric Field of View. *The International Journal of Aviation Psychology, 5*(3), 233-256.

Bechara, A., Damasio, A., Damasio, H., and Anderson, S. (1994). Insensitvity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7-15.

Bechara, A., Damasio, A., Tranel, D., and Anderson, S. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *The Journal of Neuroscience, 18*(1), 428-437.

Bechara, A., Tranel, D., and Damasio, A. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain, 123*(11), 2189-2202.

Belenky, G. (1997). *Sleep, Sleep Deprivation, and Human Performance in Continuous Operations*, from http://www.usafa.af.mil/jscope/JSCOPE97/Belenky97/Belenky97.htm

Belenky, G., Penetar, D., Throne, D., Popp, K., Leu, J., Thomas, M., et al. (1994). The effects of sleep deprivation on performance during continuous combat operations. In B. M. Marriot (Ed.), *Food Components to Enhance Performance* (pp. 127-135). Washington D.C.: National Academy Press.

Belenky, G., Wesensten, N., Thorne, D., Thomas, M., Sing, H., Redmond, D., et al. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research, 12*, 1-12.

Bennington, J. (2000). Sleep Homeostasis and the Function of Sleep. *Sleep, 23*(7), 959-966.

Boag, C. (2003).

Booher, H. (2003). Introduction: Human Systems Integration. In H. Booher (Ed.), *Handbook of Human Systems Integration* (pp. 1-30): John Wiley and Sons.

Boulay, L. (2004). Contrasting the effects of haloperidol and olanzapine on attention and working memory in schizophrenia: a double blind flexible dose study. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 64*(9-B), 4659.

Brown, E. (2004). *The Effects of Fatigue on Position Determination and Cognitive Workload Using a Visual and 3-Dimensional Auditory Display*. Naval Postgraduate School, Monterey, CA.

Caldwell, J., and LeDuc, P. (1998). Gender influences on performance, mood, and recovery sleep in fatigued aviators. *Ergonomics, 41*(12), 1757-1170.

Caldwell, J. A. (1997). Fatigue in the aviation environment: an overview of the causes and effects as well as recommended countermeasures. *Aviation, Space, and Environmental Medicine, 68*(10), 932-938.

Caldwell, J. A., and Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of U.S. Army aviation personnel. *Aviation, Space, and Environmental Medicine, 73*(5), 472-479.

Carretta, T. R., Perry, D. C., and Ree, M. J. (1996). Prediction of Situational Awareness in F-15 Pilots. *International Journal of Aviation Psychology, 6*(1), 21-41.

Carretta, T. R., and Ree, M. J. (2003). Pilot Selection Methods. In P. S. Tsang and M. A. Vidulich (Eds.), *Principles and Practice in Aviation Psychology* (pp. 357-396). Mahwah, NJ: Lawrence Erlbaum Associates.

Carter, R. (1998). *Mapping the Mind*. Berkeley: University of California Press.

Chee, M., and Choo, W. (2004). Functional imaging of working memory after 24 hr of total sleep deprivation. *Journal of Neuroscience, 24*(19), 4560-4567.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., et al. (2002). Driver sleepiness and risk of serious injury to car occupants: population based control study. *British Journal of Medicine, 324*, 1125-1129.

Cooper, R. (2002). *Modeling High-Level Cognitive Processes*. Mahwah: Lawrence Erlbaum Associates.

Crowley, J., and Geyer, S. L. (1993). Helicopter rotor blade injury: a persistent safety hazard in the U.S. Army. *Aviation, Space, and Environmental Medicine, 64*(9 Pt 1), 854-858.

Damos, D. L. (1996). Pilot Selection Batteries: Shortcomings and Perspectives. *International Journal of Aviation Psychology, 6*(2), 199-209.

De Valck, E., and Cluydts, R. (2001). Slow-release caffeine as a countermeasure to driver sleepiness induced by partial sleep deprivation. *Journal of Sleep Research, 10*, 203-209.

D'Esposito, M., Detre, J., Alsop, D., Shin, R., Atlas, S., and Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*, 279.

The development of situation awareness measures in ATM systems. (No. HRS/HSP-005-REP-01)(2003). European Organisation for the Safety of Air Navigation.

Dinges, D. F. (1995). Sleep Deprivation and Human Immune Function. *Advances in Neuroimmunology, 5*, 97-110.

Dinges, D. F. (2001). Stress, Fatigue, and Behavioral Energy. *Nutrition Reviews, 59*(1), S30-S32.

Donders, J., Tulsky, D., and Zhu, J. (2001). Criterion validity of new WAIS-III subtest scores after traumatic brain injury. *Journal of the International Neuropsychological Society, 7*(7), 892-989.

Drummond, S., Brown, G., Gillin, C., Stricker, J., Wong, E., and Buxton, R. (2000). Altered brain response to verbal learning following sleep deprivation. *Nature, 403*(6770), 655-657.

Drummond, S., Brown, G., Stricker, J., Buxton, R., Wong, E., and Gillin, C. (1999). Sleep deprivation-induced reduction in cortical functional response to serial subtraction. *NeuroReport, 10*, 3745-3748.

Drummond, S., Gillin, C., and Brown, G. (2001). Increased cerebral response during a divided attention task following sleep deprivation. *Journal of Sleep Research, 10*, 85-92.

Drummond, S. P. A., and Brown, G. G. (2001). The Effects of Total Sleep Deprivation on Cerebral Responses to Cognitive Performance. *Neuropsychopharmacology, 25*(55), S68-S73.

Durso, F., and Gronlund, S. (1999). Situation Awareness. In F. Durso (Ed.), *Handbook of Applied Cognition*. Chichester: John Wiley and Sons.

Ekstrom, R., French, J., and Harman, H. (1976). *Manual for the Kit of Factor-Referenced Cognitive Tests*. Princetown, NJ: Educational Testing Services.

Endsley, M. R. (1988). *Design and evaluation for situation awareness enhancement.* Paper presented at the Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA.

Endsley, M. R. (1988). *Situation Awareness Global Assessment Technique.* Paper presented at the Proceedings of the National Aerospace and Electronics Conference (NAECON), New York.

Endsley, M. R. (1990a). A Methodology for the Objective Measurement of Pilot Situation Awareness. In *Situation Awareness in Aerospace Operations (AGARD-CP-478)* (pp. 1/1-1/9). Neuilly Sur Seine, France: NATO-AGARD.

Endsley, M. R. (1990b). *Predictive utility of an objective measure of situation awareness.* Paper presented at the Proceedings of the Human Factors Society 34th Annual Meeting.

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (1995b). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (1999). Situation Awareness in Aviation Systems. In D. Garland, J. Wise and V. Hopkins (Eds.), *Handbook of Aviation Human Factors*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000a). Direct Measurement of Situation Awareness: Validity and Use of SAGAT. In M. R. Endsley and D. Garland (Eds.), *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000b). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley and D. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Ernst, M., Bolla, K., Mouratidis, M., Contoreggi, C., Matochik, J., Kurian, V., et al. (2002). Decision-making in a risk-taking task: a PET study. *Neuropsychopharmacology, 26*(5), 682-691.

Ferrara, M., and DeGennaro, L. (2001). How much sleep do we need? *Physiological Review, 5*(2), 155-179.

Flach, J. (1995). Situation awareness: proceed with caution. *Human Factors, 37*(1), 149-157.

Frequently Asked Questions. (2003). from http://www.aero.ca/e_faq.html

Friedmann, J., Globus, G., Huntley, A., Mullaney, D., Naitoh, P., and Johnson, L. (1977). Performance and mood during and after gradual sleep reduction. *Psychophysiology, 14*(3), 245-250.

Frith, C., and Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. *Cognitive Brain Research, 5*, 175-181.

The Functions of Sleep. from http://www.portfolio.mvm.ed.ac.uk/studentwebs/session2/group11/functions.htm

Fuster, J. (1989). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* (2nd ed.). New York: Raven Press.

Garson, D. (2005). *Data Levels and Measurement*, from http://www2.chass.ncsu.edu/garson/pa765/datalevl.htm

Giam, G. (1997). Effects of Sleep Deprivation with Reference to Military Operations. *Annals of Academic Medicine Singapore, 26*, 88-93.

Glenville, M., Broughton, R., Wing, A., and Wilkinson, R. (1978). Effects of sleep deprivation on short duration performance measures compared to the Wilkinson auditory vigilance task. *Sleep, 1*(2), 169-176.

Goel, V., and Dolan, R. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition, 93*(3), B109-121.

Goldberg, E. (2001). *The Executive Brain: Frontal lobes and the civilized mind*. New York: Oxford University Press.

Gosselin, A., DeKoninck, J., and Campbell, K. (2005). Total sleep deprivation and novelty processing: implications for frontal lobe functioning. *Clinical Neurophysiology, 116*(1), 211-222.

Hanford, R. (2001). Working memory and nonverbal affect recognition: differentiation among ADHD subtypes in adults. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 61*(10-B), 5564.

Harma, M. (1995). Sleepiness and shiftwork: individual differences. *Journal of Sleep Research, 4*(Suppl 2), 57-61.

Harrison, Y., and Horne, J. (1998). Sleep loss impairs short and novel language tasks having a prefrontal focus. *Journal of Sleep Research, 7*, 95-100.

Harrison, Y., and Horne, J. (1999). One night of sleep loss impairs innovative thinking and flexible decision making. *Organizational Behavior and Human Decision Processes, 78*(2), 128-145.

Hartman, B., and Secrist, G. (1991). Situational awareness is more than exceptional vision. *Aviation, Space, and Environmental Medicine, 62*(11), 1084-1089.

Hartman, B. O., and Secrist, G. E. (1991). Situational Awareness is More than Exceptional Vision. *Aviation, Space, and Environmental Medicine, 62*(11), 1084-1089.

Haut, M., Kuwabara, H., Leach, S., and Arias, R. (2000). Neural Activation During Performance of Number-Letter Sequencing. *Applied Neuropsychology, 7*, 237-242.

Herscovitch, J., Stuss, D., and Broughton, R. (1980). Changes in cognitive processing following short-term cumulative partial sleep deprivation and recovery oversleeping. *Journal of Clinical Neuropsychology, 2*, 301-319.

Himashree, G., Banerjee, P. K., and Selvamurthy, W. (2002). Sleep and Performance-Recent Trends. *Indian Journal of Physiology and Pharmacology, 46*(1), 6-24.

Hockey, G., Wastell, D., and Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: A multilevel analysis of compensatory control. *Human Factors, 40*(2), 233-253.

Hollander, M., and Wolfe, D. (1999). *Nonparametric Statistical Methods* (2 ed.). New York: John Wiley and Sons.

Hopkin, D. (1993). *Situation awareness in air traffic control.* Paper presented at the Situational Awareness in Complex Systems Conference, Orlando, FL.

Horne, J. (1988). Sleep loss and "divergent" thinking ability. *Sleep, 11*(6), 528-536.

Horne, J. (1993). Human sleep, sleep loss and behaviour: Implications for the prefrontal cortex and psychiatric disorder. *British Journal of Psychiatry, 162*, 413-419.

Horne, J., and Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*(2), 97-110.

How to Perform and Interpret Factor Analysis using SPSS. (2005). from http://www.ncl.ac.uk/iss/statistics/docs/factoranalysis.html

Introduction to brain imaging techniques and other methods. (2003). from http://psy.st-andrews.ac.uk/resources/brain.shtml

Jay, C. (2003, July 3, 2003). US Army tackles sleep deprivation. *Australian Financial Review*.

Jones, D. G., and Endsley, M. R. (1996). Sources of Situation Awareness Errors in Aviation. *Aviation, Space, and Environmental Medicine, 67*(6), 507-512.

Jones, K., and Harrison, Y. (2001). Frontal lobe function, sleep loss, and fragmented sleep. *Sleep Medicine Reviews, 5*(6), 463-475.

Karnath, H., and Wallesch, C. (1992). Inflexibility of mental planning: a characteristic disorder with prefrontal lobe lesions? *Neuropsychologia, 30*(11), 1011-1016.

Karnath, H., Wallesch, C., and Zimmermann, P. (1991). Mental planning and anticipatory processes with acute and chronic frontal lobe lesions: a comparison of maze performance in routine and non-routine situations. *Neuropsychologia, 29*(4), 271-290.

Kent, A., Fox, A., Michie, P., and Jablensky, A. (2004). Differential impairment of working memory performance in first-degree relatives of individuals with schizophrenia. *Acta Neuropsychiatrica, 16*(3), 149-153.

Kerkhof, G. A. (1985). Inter-individual Differences in the Human Circadian System: A Review. *Biological Psychology, 20*, 83-112.

Killgore, W., Balkin, T., and Wesensten, N. (in submission). Impaired decision-making following 49 hours of sleep deprivation. *Journal of Sleep Research*.

Killgore, D. (2004). Personal Communication.

Kleiner, B. M., and Booher, H. (2003). Human Systems Integration Education and Training. In H. Booher (Ed.), *Handbook of Human Systems Integration* (pp. 121-164): John Wiley and Sons.

Knapp, C. J., and Johnson, R. (1996). F-16 Class A Mishaps in the U.S. Air Force. *Aviation, Space, and Environmental Medicine, 67*(8), 777-783.

Kolb, B., and Whishaw, I. (1996). *Fundamentals of Human Neuropsychology* (4th ed.). New York: W.H. Freeman and Company.

Kolb, B., and Wishaw, I. (1985). *Fundamentals of Human Neuropsychology* (2nd ed.). San Francisco: W.H. Freeman.

Krawczyk, D. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neuroscience and Biobehavioral Reviews, 26*, 631-664.

Kreuger, J., Obal, F., and Fang, J. (1999). Why we sleep: a theoretical view of sleep function. *Sleep Medicine Reviews, 3*(2), 119-129.

Langdon, H.W. and Merino, B.J. (1992). Acquisition and development of a second language in the Spanish speaker, In H.W. Langdon and L.L. Cheng, *Hispanic children and adults with communication disorders*, (pp.132-167). Gaithersburg, MD: Aspen.

LeDuc, P., Riley, D., Hoffman, S. M., Brock, M. E., Norman, D., Johnson, P., et al. (1999). *The Effects of Sleep Deprivation on Spatial Disorientation* (No. USAARL 2000-09). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Lee, J., Dingus, T., Mollenhauer, M., Brown, T., and Neale, V. (1997). *Development of human factors guidelines for advanced traveler information systems (ATIS) and commercial*

vehicle operations (CVO): CVO driver fatigue and complex in-vehicle systems (No. FHWA-RD-96-151): Federal Highway Institution.

Lieberman, A. R., Tharion, W. J., Shukitt-Hale, B., Speckman, K. L., and Tulley, R. (2002). Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U.S. Navy SEAL training. *Psychopharmacology, 164*, 250-261.

Linde, L., and Bergstrom, M. (1992). The effect of one night without sleep on problem-solving and immediate recall. *Psychological Research, 54*(2), 127-136.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain, 125*, 624-639.

Marsh, H. (2000). *Beyond situation awareness: The battlespace of the future*: Office of Naval Research.

Martin, G. (2003). *Essential Biological Psychology*. London: Arnold.

Mathias, R. (1996). The basics of brain imaging. *NIDA Notes, 11*.

May, J., and Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *British Journal of Psychology, 78*, 443-455.

McGurk, S., T, C., Harvey, P., Reichenberg, A., White, L., Friedman, J., et al. (2004). Working memory performance in poor outcome schizophrenia: relationship to age and executive functioning. *Journal of Clinical and Experimental Neuropsychology, 26*(2), 153-160.

Meerlo, P., Koehl, M., van dar Borght, K., and Turek, F. (2002). Sleep restriction alters the hypothalamic-pituitary-adrenal response to stress. *Journal of Neuroendocrinology, 14*, 397-402.

Montgomery, D. (1997). *Design and Analysis of Experiments* (4 ed.). New York: John Wiley and Sons.

Moore-Ede, M. (Writer) (2003). Sleep, alertness, and fatigue among military personnel in Iraq. In I. Circadian Technologies (Producer). Lexington, MA.

Muzur, A., Pace-Schott, E., and Hobson, J. (2002). The prefrontal cortex in sleep. *Trends in Cognitive Sciences, 6*(11), 475-481.

Need for speed: Did amphetamines play a role in Afghanistan friendly fire incident? (2002). Retrieved April, 2003, from http://abcnews.go.com/sections/2020/DailyNews/2020_pilotpills021220.html

Nielson, T. (2000). Mentation in REM and NREM Sleep: A review and possible reconciliation of two models. *Behavioral and Brain Sciences, 23*(6), ???

Nindl, B., Leone, C., Tharion, W., Johnson, R., Castellani, J., Patton, J., et al. (2002). Physical performance responses during 72 h of military operational stress. *Medicine and Science in Sports and Exercise, 34*(11), 1814-1822.

O'Hare, D. (1997). Cognitive Ability Determinants of Elite Pilot Performance. *Human Factors, 39*(4), 540-552.

Partinen, M. (1994). Sleep Disorders and Stress. *Journal of Psychosomatic Research, 38*(Suppl 1), 89-91.

Peters, R. D., Wagner, E., Alicandri, E., Fox, J. E., Thomas, M., Thorne, D., et al. (1999). Effects of Partial and Total Sleep Deprivation on Driving Performance. *Public Roads, 62*(4).

Peterson, S., van Mier, H., Fiez, J., and Raichle, M. (1998). The effects of practice on the functional anatomy of task performance. *Proceedings of the National Academy of Sciences, 95*(3), 853-860.

Randel, J., and Pugh, L. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Interaction, 45*, 579-597.

Rechtschaffen, A., and Kales, A. (1968). *A manual of standardised terminology, techniques and scoring system for sleep stages of human subjects*. Los Angeles: Brain Information service/Brain Research Institute.

Roscoe, S. (1997). Predicting and enhancing flightdeck performance. In R. Telfer and P. Moore (Eds.), *Aviation Training: Learners, Instruction, and Organization* (pp. 195-208). Aldershot, England: Avebury.

Roscoe, S. (2003). Personal communication.

Roscoe, S., Corl, L., and LaRoche, J. (2001). *Predicting Human Performance* (5th ed.): Helio Press.

Ryan, J. J., and Paolo, A. M. (2001). Exploratory factor analysis of the WAIS-III in a mixed patient sample. *Arch Clin Neuropsychol, 16*(2), 151-156.

Ryman, D., Naitoh, P., and Englund, C. (1985). Decrements in logical reasoning performance under conditions of sleep loss and physical exercise: the factor of sentence complexity. *Perceptual and Motor Skills, 61*(3 Part 2), 1179-1188.

Sanders, M., and McCormick, E. (1993). *Human Factors in Engineering and Design* (7 ed.). New York: McGraw-Hill.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., and Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*, 1755-1758.

Sarter, N., and Woods, D. D. (1991). Situation Awareness: A Critical But Ill-Defined Phenomenon. *International Journal of Aviation Psychology, 1*(1), 45-57.

Selcon, S. J., and Taylor, R. (1989). Evaluation of the Situational Awareness Rating Technique (SART) as a Tool for Aircrew Systems Design. In *Situational Awareness in aerospace Operations (AGARD-CP-478)*. Neuilly Sur Seine, France: NATO-AGARD.

Smith, K., and Hancock, P. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors, 37*(1), 137-148.

Steinweg, K. K. (1995). Dealing realistically with fratricide. *Parameters*(Spring), 4-29.

Subbotsky, E.*Vigotsky's distinction between lower and higher mental functions and recent studies on infant cognitive development*, from http://psych.hanover.edu/vygotsky/subbot.html

Taylor, R. (1990). Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design. In *Situational Awareness in Aerospace Operations (AGARD-CP-478)* (pp. 3/1-3/17). Neuilly Sur Seine, France: NATO-AGARD.

Technology, N. I. o. S. a. (2005). *NIST/SEMATECH e-Handbook of Statistical Methods*, from http://www.itl.nist.gov/div898/handbook/2005

Thomas, M., Sing, H., Belenky, G., Holcomb, H., Mayberg, H., Dannals, R., et al. (2000). Neural basis of alertness and cognitive performance impairments during sleepiness. I. Effects of 24 h of sleep deprivation on waking human regional brain activity. *Journal of Sleep Research, 9*, 335-352.

Thomas, M., Sing, H., Belenky, G., Holcomb, H., Mayberg, h., Dannals, R., et al. (2000). neural basis of alertness and cognitive impairments during sleepiness. I. Effects of 24h of sleep deprivation on waking human regional brain activity. *Journal of Sleep Research, 9*, 335-352.

Uhlarik, J., and Comerford, D. (2002). A Review of Situation Awareness Literature Relevant to Pilot Surveillance Functions. *FAA Office of Aviation Medicine Reports, DOT/FAA/AM-02/3*, 24.

Van Dongen, H., Baynard, M., Maislin, G., and Dinges, D. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep, 27*(3), 423-433.

Vihman, M.M. and McLaughlin, B. (1982). Bilingualism and second language acquisition in preschool children. In C.J. Brainerd and M. Pressley (Eds.), *Verbal processes in children: Progress in cognitive development research* (pp.35-57). New York: Spring Verlag.

Vincent, K. (1991). Black/white IQ differences: does age make the difference? *Journal of Clinical Psychology, 47*(2), 266-270.

The Way Ahead. (2003). Washington, DC: Army Strategic Communication.

Wechsler, D. *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wesensten, N. (2003). *Effects of single-dose caffeine, dextroamphetamine, or modafinil on performance and objective alertness during sleep deprivation*. Silver Spring, MD: Walter Reed Army Institute of Research.

Wickens, C. D. (2002a). Situation Awareness and Workload in Aviation. *Current Directions in Psychological Science, 11*(4), 128-133.

Wickens, C. D. (2002b). *Spatial Awareness Biases* (No. ARL-02-6/NASA-02-4). Savoy, IL: Aviation Research Lab Institute of Aviation.

Wickens, C. D., and Hollands, J. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River: Prentice-Hall.

Wiegmann, D. A., Stanny, R. R., McKay, D. L., Neri, D. F., and McCardie, A. H. (1996). Methamphetamine effects on cognitive processing during extended wakefulness. *The International Journal of Aviation Psychology, 6*(4), 379-397.

Wimmer, F., Hoffmann, R., Bonato, R., and Moffitt, A. (1992). The effects of sleep deprivation on divergent thinking and attention processes. *Journal of Sleep Research, 1*, 223-230.

WOMBAT: Modern Psychological Assessment. from http://www.aero.ca.e_W-Presentation.html

Wymer, J., Rayls, K., and Wagner, M. (2003). Utility of a clinically derived abbreviated form of the WAIS-III. *Archives of Clinical Neuropsychology, 18*(8), 917-927.

Zysset, S., Huber, O., Ferstl, E., and Yves von Cramon, D. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage, 15*, 983-991.

APPENDIX A. STUDY SCHEDULE

WEDNESDAY –DAY 1

WED	**1700**		Set up CATs Computer Tests (enter Ss demographic data on each computer)	
WED			White noise ON	
WED			White noise dB = 65	
WED	**1800**		Subjects report to laboratory	
WED			DAY 1 BRIEFING (verbal instructions attached at back of schedule)	
WED			Collect Sleep Diary – put in subject folder	
WED			Collect SUPERVISOR APPROVAL form (active duty only)	
WED			Collect/verify LEAVE form (active duty only)	
WED			Ss must stop chewing gum	
WED			Ss turn cell phones OFF	
WED			Ss remove pagers and/or watches	
WED			Vital Signs (BP, Pulse and tympanic temperature)	BP
				Pulse
				T Temp
WED			Urine Sample (Have subject verify his/her INITIALS, DOB, DATE on cup label)	
WED			**FEMALES: PREGNANCY TEST** (use urine from original collection)	
WED			**FEMALES: Send pregnancy test to QUEST STAT (place in red bag)**	
WED			ALL VOLUNTEERS: Send urine drug screens	
WED	**1820**		Electrode placement (2 subjects @ a time) – check electrode impedance manually	
WED			WASI Administration (1 subject @ a time, 45 min each)	
WED	**1930**		Demonstrate: MWT (see instruction booklet)	
WED			Demonstrate: Voice Task (see instruction booklet)	
WED			Demonstrate: PVT (see instruction booklet)	
WED			Demonstrate: Sleep Scales, Line Bisection, and VAMS	
WED			Computer monitors OFF	
WED	**1940**		Verify electrode attachment to SIESTA	
WED			Check electrode impedances on laptop	
WED			Check electrode recordings on laptop **- DO NOT start recording on SIESTA!**	

WED	**1950**		Day 1 Medical Screen
WED			Laterality Inventory
WED			WOMBAT: Reading instructions on Computers (up to 50 minutes)
WED			WOMBAT: Practice Session (60 minutes)
WED	**2220**		Begin Checking Electrodes – gel + impedance check + repair
WED	**2230**		**Start SIESTA RECORDING - DISK #1**
WED	**2240**		Notify subjects that bedtime is in 20 minutes
WED			Re-check electrode recordings on laptop
WED	**2255**		Ss into chamber for bedtime prep:
WED			SAY: **"I will turn off the bedroom lights in a few minutes. I will wake you up tomorrow morning at 7:00 am by turning on the lights, entering the bedroom, and announcing the time. The light will be on in the hallway all night. If you have to use the bathroom, please be quiet so you don't wake the other subjects. A technician will be sleeping in the lounge all night if you need anything."**
WED	**2315**		Shut doors / Lights out
WED			Clean lounge/kitchen area
WED			Clean examining room and put away all chemicals
WED			12 NEW "AA" batteries / 4 formatted flash disks
WED			**INSTRUCTIONS TO OVERNIGHT TECHNICIAN**: **You must stay within earshot of subjects at all times**. • Do not leave subjects unattended for any reason. • Subjects may not leave the building, have guests, or use the telephone for personal reasons at any time. • Keep TV volume low. Stereo may not be used. • If you are unsure about anything, or a situation arises, immediately call:

THURSDAY – DAY 2

THU	**0645**		**TECH SIGN-IN:** _____ _____ _____ _____

THU	**0700**		**Awaken Subjects**: • Open bedroom doors. • Say: **"It is now 7 A.M. Please come into the lounge area."** • Turn lights on AFTER you wake up subjects.
THU	**0705**		Use restroom; Vital signs (BP and pulse) BP / Pulse
THU			Electrodes: gel + impedance check + repair
THU			SIESTA: battery + flashcard replacement **DISK #2**
THU			Meal
THU	**0800**		**"In about 15 minutes, we will begin the tests you practiced last night. Throughout the study, please do not talk with each other about how you did on the tests, or how you are feeling. Discussing these things is grounds for removal from the study. On Sunday at the end of the study, you will be able to talk about all of these things with each other, and the principal investigator will answer any questions you have.**
THU	**0805**	**T1**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**0820**	**P1**	PVT (all Ss simultaneously)
THU		**P1**	Sleep Scales, Line Bisection, VAMS (escort subjects back to lounge)
THU	**0855**		Synch SIESTAs and laptops
THU			Turn on sound machines
THU	**0900**	**M1**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**1000**	**T2**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**1020**	**P2**	PVT (all Ss simultaneously)
THU		**P2**	Sleep Scales, Line Bisection, VAMS (escort subjects back to lounge)
THU	**1030**	**V1**	Voice Task
THU	**1045**		Remind subjects: "THE FIRST TESTING BLOCK WILL START IN ABOUT 45 MINUTES, FOLLOWING THE NEXT MWT. BE SURE YOU HAVE HAD A SNACK AND GONE TO THE RESTROOM BEFORE THE TESTING STARTS."
THU	**1055**		Synch SIESTAs and laptops
THU	**1100**	**M2**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**1115**		Remind subjects: "THE FIRST TESTING BLOCK WILL START IN 15 MINUTES. BE SURE YOU HAVE HAD A SNACK IF YOU ARE HUNGRY AND HAVE HAD A CHANCE TO GO TO THE RESTROOM BEFORE THE TESTING STARTS."
THU	**1120**		**Turn off sound machines**

THU	**1130**	**1**	**EKMAN 60**
THU	**1140**	**1**	**JLO** (**FORM H**)
THU	**1150**	**1**	**Iowa Gambling Task (IGT)** Load: **ABCD**XPGTBPSP.EXE
THU		**1**	Rename Gambling Task "Output file" as "IGT_Subjectnumber_1.txt" and drag it to the DATA folder on the desktop
THU	**1210**	**T3**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**1212**	**1**	**EVAR and BSSS (FORM A)**
THU	**1220**	**P3**	PVT (all Ss simultaneously)
THU		**P3**	Sleep Scales, Line Bisection, VAMS
THU	**1230**	**1**	**EHT (Emotion Hexagon Test)**
THU	**1250**	**1**	**BART**
THU	**1255**		Synch SIESTAs and laptops
THU			Turn on sound machines
THU			Turn off monitors in bedrooms
THU	**1310**	**M3**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**1325**	**1**	**UPSIT** (Smell Test—**FORM A**)
THU	**1335**	**1**	**STROOP**
THU	**1345**	**1**	**COWA-verbal fluency** (**CFL**) and **ANIMALS**
THU	**1355**	**1**	**COLOR TRAILS** (**FORM A**)
THU	**1405**	**T4**	**LEFT EAR** Tympanic Temperature (record in blank)
THU			Snack / Meal Break (if time permits)
THU	**1420**	**P4**	PVT (all Ss simultaneously)
THU		**P4**	Sleep Scales, Line Bisection, VAMS
THU	**1430**	**1**	**MAC** Tests—**Emotion Differentiation Test**
THU		**1**	**MAC** Tests—**Emotion Acuity Test**
THU			SIESTA - battery + flashcard replacement **DISK #3**
THU	**1445**		**TECH SIGN-IN:** _____ _____ _____ _____

THU	**1455**		Synch SIESTAs and laptops
THU			Turn on sound machines
THU	**1500**	**M4**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**1515**		Remind subjects: "THE NEXT TESTING BLOCK WILL START IN 5 MINUTES. IF YOU NEED TO USE THE RESTROOM, PLEASE DO SO NOW AND RETURN TO YOUR ROOM WITHIN THE NEXT 5 MINUTES."
THU			Turn off sound machines
THU	**1520**	**1**	**WOMBAT**
THU	**1620**	**P5**	PVT (all Ss simultaneously)
		P5	Sleep Scales, Line Bisection, VAMS
THU		**T5**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**1630**	**1**	Mazes (**Form 1**) and Letter Number Sequencing (**Form A**)
THU	**1640**	**1**	Letter Sets (**Form 1**)
THU	**1650**	**1**	Nonsense Syllogisms (**Form 1**)
THU			Synch SIESTAs and laptops
THU			Turn on sound machines
THU			Turn off monitors in bedrooms
THU	**1700**	**M5**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**1720**		Meal
THU	**1730**		Vital signs (BP and pulse) BP / Pulse
THU			Electrodes – gel + impedance check + repair
THU			Urine Sample (Have subject verify his/her INITIALS Subject number, DOB, DATE on cup label)
THU			ALL VOLUNTEERS: Send urine drug screens
THU	**1800**	**T6**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**1810**		Check Quest Printer in blood lab to see if Pregnancy Screens have come in. VERIFY THAT PREGNANCY IS NEGATIVE and place the report in subject's medical file. IF TEST IS POSITIVE, NOTIFY PI!!!
THU	**1820**	**P6**	PVT (all Ss simultaneously)
THU		**P6**	Sleep Scales, Line Bisection, VAMS

THU	**1830**	**V2**	Voice Task
THU	**1850**		Synch SIESTAs and laptops
THU	**1900**	**M6**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**2000**	**T7**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**2020**	**P7**	PVT (all Ss simultaneously)
		P7	Sleep Scales, Line Bisection, VAMS
THU	**2050**		Synch SIESTAs and laptops
THU	**2100**	**M7**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**2200**	**T8**	**LEFT EAR** Tympanic Temperature (record in blank)
THU	**2220**	**P8**	PVT (all Ss simultaneously)
THU		**P8**	Sleep Scales, Line Bisection, VAMS
THU	**2230**	**V3**	Voice Task
THU	**2235**		SIESTA - battery + flashcard replacement **DISK #4**
THU	**2245**		**TECH SIGN-IN:** _____ _____ _____ _____
THU	**2250**		Synch SIESTAs and laptops
THU	**2300**	**M8**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
THU	**2320**		Meal
THU	**2330**		Vital signs (BP and pulse) BP / Pulse
THU			Electrodes – gel + impedance check + repair

FRIDAY – DAY 3

FRI	**0000**	**T9**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI	**0020**	**P9**	PVT (all Ss simultaneously)

FRI		**P9**	Sleep Scales, Line Bisection, VAMS
FRI	**0030**	**V4**	Voice Task
FRI	**0050**		Synch SIESTAs and laptops
FRI	**0100**	**M9**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
FRI	**0200**	**T10**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI	**0220**	**P10**	PVT (all Ss simultaneously)
FRI		**P10**	Sleep Scales, Line Bisection, VAMS
FRI	**0230**	**V5**	Voice Task
FRI	**0250**		Synch SIESTAs and laptops
FRI	**0300**	**M10**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
FRI	**0400**	**T11**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI	**0420**	**P11**	PVT (all Ss simultaneously)
FRI		**P11**	Sleep Scales, Line Bisection, VAMS
FRI	**0430**	**V6**	Voice Task
FRI	**0435**		Electrodes: gel and impedance check
FRI			Remind subjects: "THE NEXT TESTING BLOCK WILL START IN ABOUT AN HOUR, FOLLOWING THE NEXT MWT. BE SURE YOU HAVE HAD A SNACK AND GONE TO THE RESTROOM BEFORE THE TESTING STARTS."
FRI	**0450**		Synch SIESTAs and laptops
FRI	**0500**	**M11**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
FRI	**0515**		Remind subjects: "THE NEXT TESTING BLOCK WILL START IN 15 MINUTES. BE SURE YOU HAVE HAD A SNACK IF YOU ARE HUNGRY AND HAVE HAD A CHANCE TO GO TO THE RESTROOM BEFORE THE TESTING STARTS."
FRI			Turn off sound machines
FRI	**0530**	**2**	**Ekman 60**
FRI	**0540**	**2**	**JLO (FORM V)**
FRI	**0550**	**2**	**Iowa Gambling Task (IGT)** Load: **KLMN**XPGTBPSP.EXE
FRI		**2**	Rename Gambling Task "Output file" as "IGT_Subjectnumber_2.txt" and drag it to the DATA folder on the desktop
FRI	**0610**	**T12**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI	**0612**	**2**	**EVAR and BSSS (FORM B)**

FRI	**0620**	**P12**	PVT (all Ss simultaneously)
FRI		**P12**	Sleep Scales, Line Bisection, VAMS
FRI	**0630**	**2**	**EHT (Emotion Hexagon Test)**
FRI	**0650**	**2**	**BART**
FRI	**0645**		**TECH SIGN-IN:** _____ _____ _____ _____
FRI	**0650**		Synch SIESTAs and laptops
FRI			Turn off monitors in bedrooms
FRI			Turn on sound machines
FRI	**0705**	**M12**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)
FRI			Turn OFF sound machines
FRI	**0725**	**2**	**UPSIT** (Smell Test—**FORM B**)
FRI	**0735**	**2**	**STROOP**
FRI	**0745**	**2**	**COWA-verbal fluency** (**PRW**) and ANIMALS
FRI	**0755**	**2**	**COLOR TRAILS (FORM B)**
FRI	**0805**	**T13**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI			Snack / Meal Break (if time permits)
FRI	**0820**	**P13**	PVT (all Ss simultaneously)
FRI		**P13**	Sleep Scales, Line Bisection, VAMS
FRI	**0830**	**2**	**MAC** Tests—**Emotion Differentiation Test**
FRI		**2**	**MAC** Tests—**Emotion Acuity Test**
FRI			SIESTA – battery + flashcard replacement **DISK #5**
FRI	**0845**		Short Break / Snack / Restroom
FRI	**0850**		Synch SIESTAs and laptops
FRI			Turn on sound machines
FRI	**0900**	**M13**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)

FRI	**0915**		Remind subjects: "THE NEXT TESTING BLOCK WILL START IN 5 MINUTES. IF YOU NEED TO USE THE RESTROOM, PLEASE DO SO NOW AND RETURN TO YOUR ROOM WITHIN THE NEXT 5 MINUTES."
FRI			Turn off sound machines
FRI	**0920**	**2**	**WOMBAT**
FRI	**1020**	**P14**	PVT (all Ss simultaneously)
FRI		**P14**	Sleep Scales, Line Bisection, VAMS
FRI		**T14**	**LEFT EAR** Tympanic Temperature (record in blank)
FRI	**1030**	**2**	Mazes (**Form 2**) and Letter Number Sequencing (**Form B**)
FRI	**1040**	**2**	Letter Sets (**Form 2**)
FRI	**1050**	**2**	NONSENSE SYLLOGISMS (**Form 2**)
FRI			Synch SIESTAs and laptops
FRI			Turn off monitors in bedrooms
FRI			Turn on sound machines
FRI	**1100**	**M14**	MWT -- Laptop 2 SYSTEM TIME: _____ (HH:MM:SS) MWT -- Laptop 3 SYSTEM TIME: _____ (HH:MM:SS)

APPENDIX B. DEMOGRAPHIC DATA

B.1 Normality Tests on Raw Data

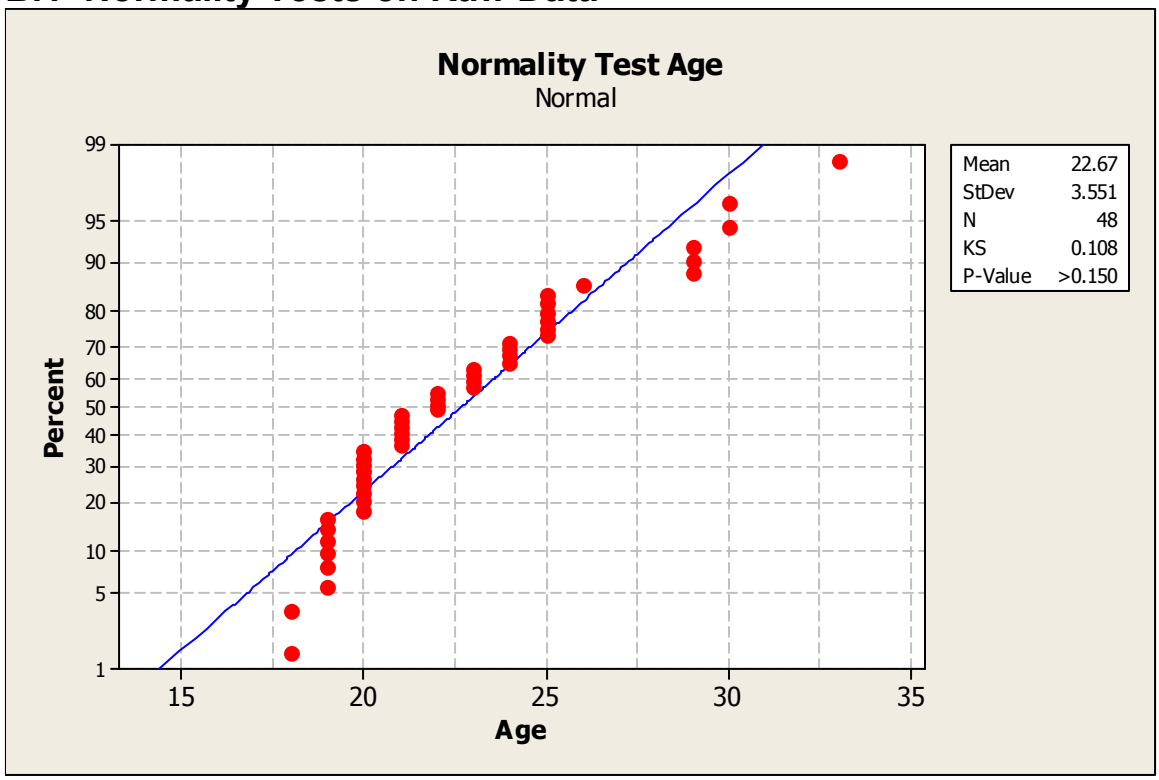


Figure B.1.1 Normality Test for Age

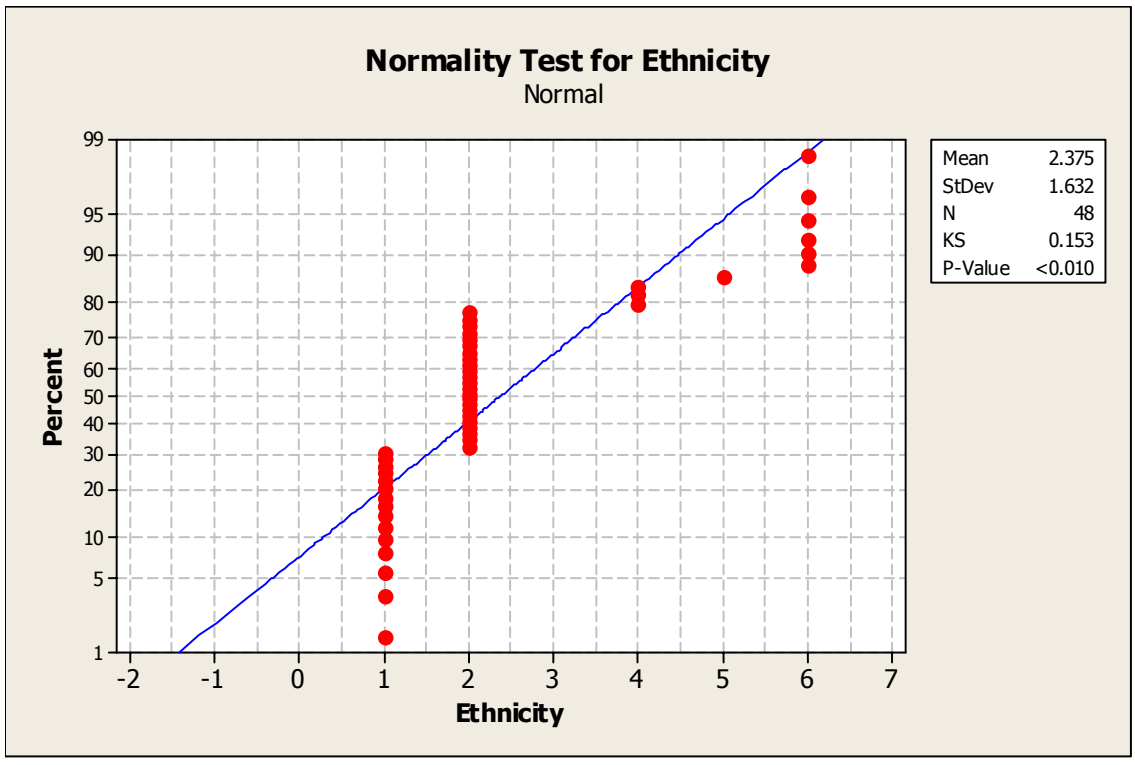


Figure B.1.2. Normality Test for Ethnicity

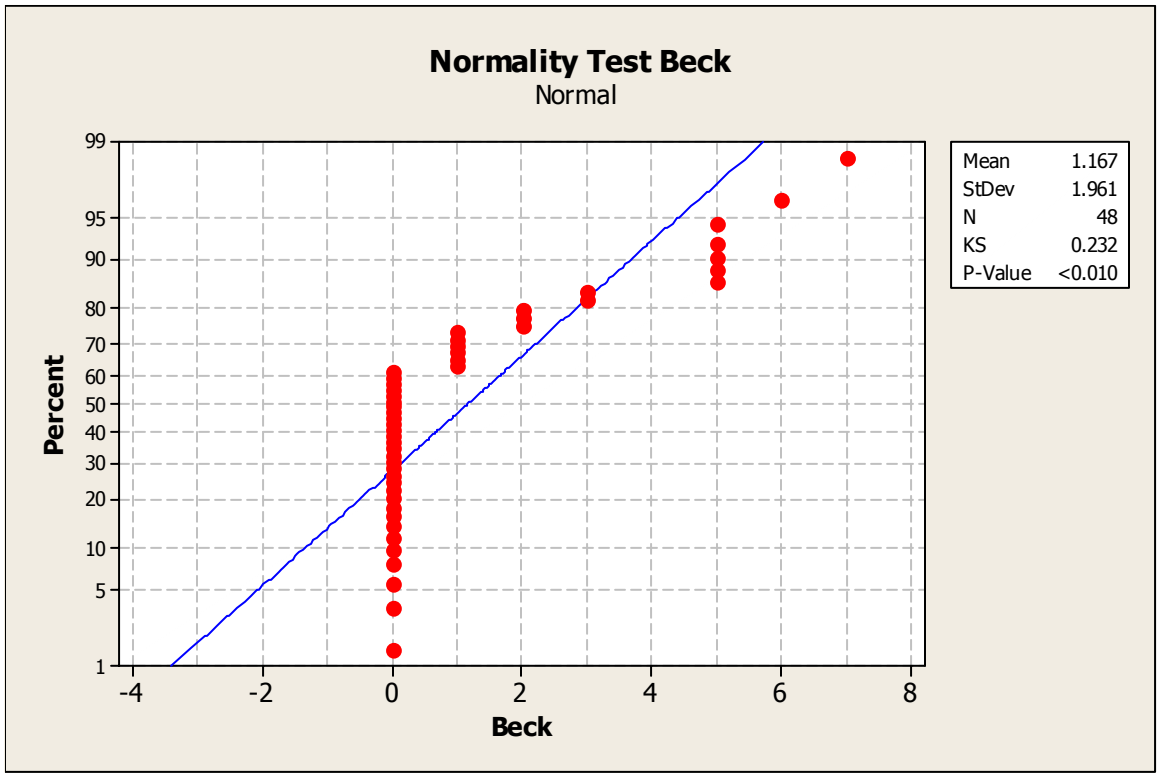


Figure B.1.3. Normality Test for Beck Depression Inventory

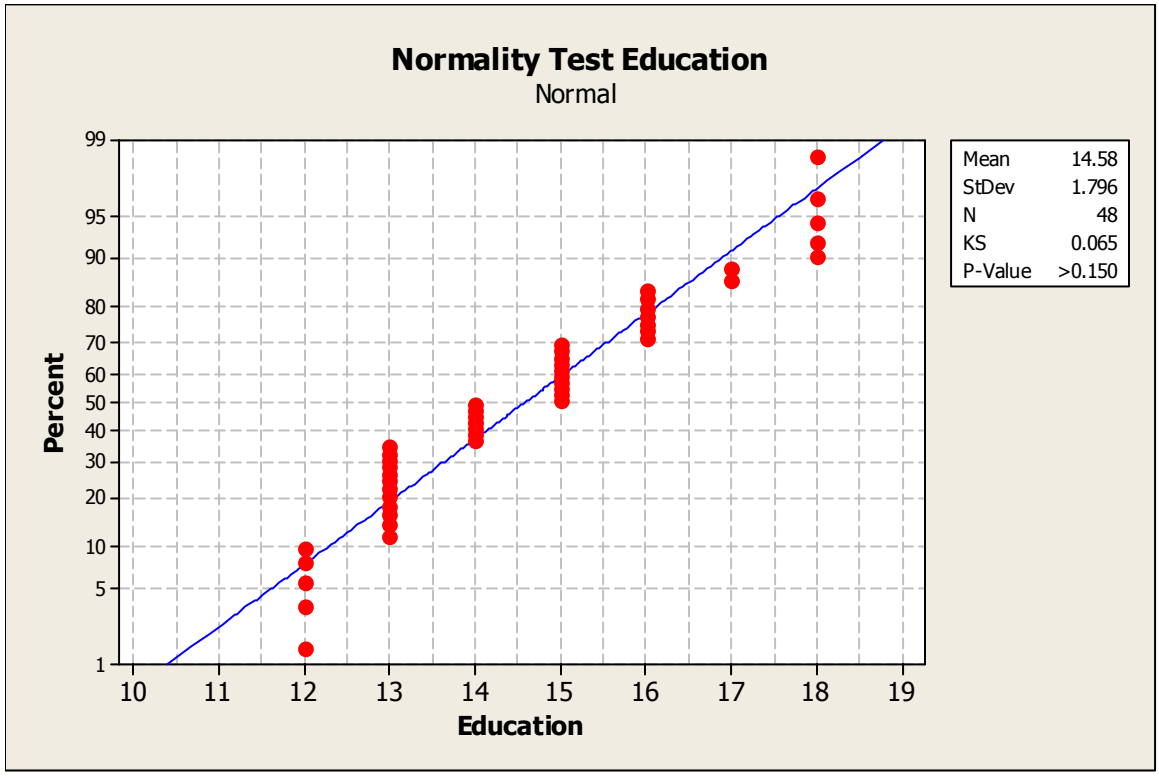


Figure B.1.4. Normality Test for Education

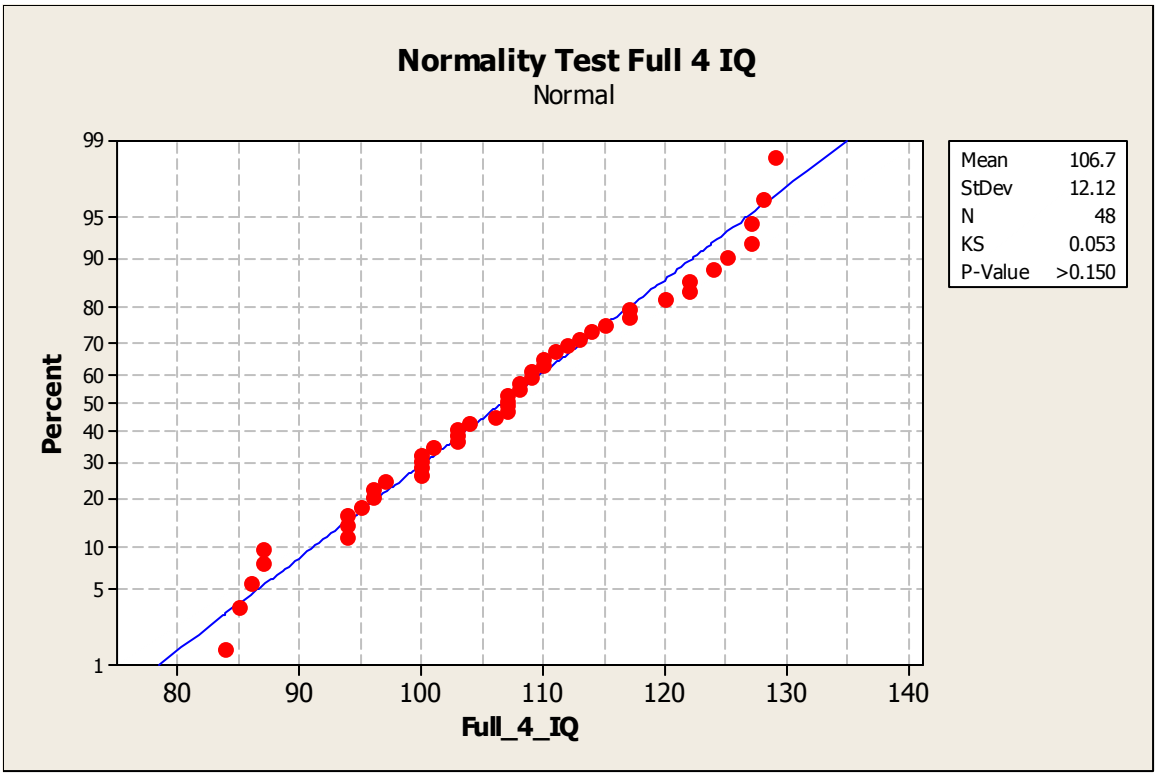


Figure B.1.5. Normality Test for WASI Full 4 IQ Score

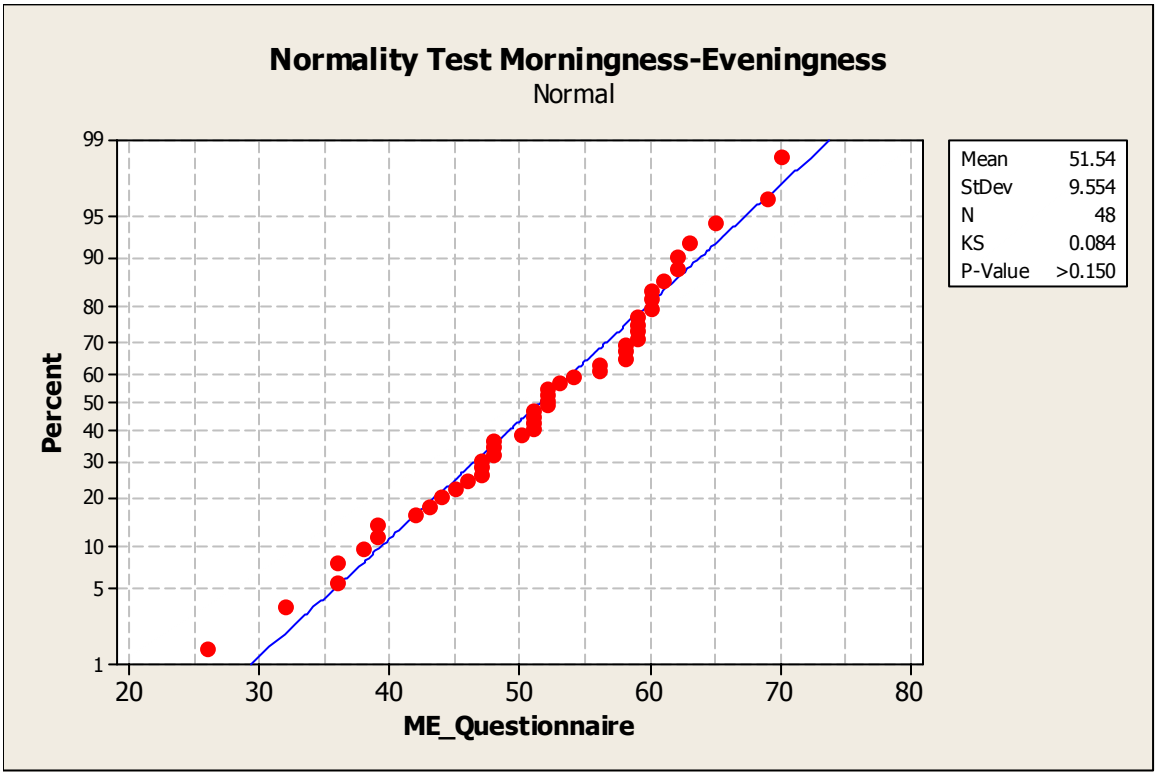


Figure B.1.6. Normality Test for Morningness-Eveningness Score

B.2 WASI Score and Demographics

Table B.2.1 ANOVA (IQ, Language)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6.420	2	3.210	.021	.979
Within Groups	6900.246	45	153.339		
Total	6906.667	47			

Table B.2.2 Tukey Post Hoc Tests (IQ, Language)

(I) Language	(J) Language	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	1.171	5.920	.979	-13.18	15.52
	3	-.104	4.853	1.000	-11.86	11.66
2	1	-1.171	5.920	.979	-15.52	13.18
	3	-1.275	7.059	.982	-18.38	15.83
3	1	.104	4.853	1.000	-11.66	11.86
	2	1.275	7.059	.982	-15.83	18.38

Table B.2.3 ANOVA (IQ, Education)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2711.938	6	451.990	4.418	.002
Within Groups	4194.729	41	102.310		
Total	6906.667	47			

Table B.2.4 Tukey Post Hoc Tests (IQ, Education)

(I) Education	(J) Education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
12	13	-15.400	5.384	.087	-32.09	1.29
	14	-22.114(*)	5.923	.010	-40.47	-3.76
	15	-23.800(*)	5.540	.002	-40.97	-6.63
	16	-7.400	5.923	.870	-25.76	10.96
	17	-9.900	8.463	.901	-36.13	16.33
	18	-14.800	6.397	.262	-34.63	5.03
13	12	15.400	5.384	.087	-1.29	32.09
	14	-6.714	4.811	.801	-21.62	8.19
	15	-8.400	4.331	.467	-21.82	5.02
	16	8.000	4.811	.643	-6.91	22.91
	17	5.500	7.725	.991	-18.44	29.44
	18	.600	5.384	1.000	-16.09	17.29
14	12	22.114(*)	5.923	.010	3.76	40.47
	13	6.714	4.811	.801	-8.19	21.62
	15	-1.686	4.985	1.000	-17.13	13.76
	16	14.714	5.407	.118	-2.04	31.47
	17	12.214	8.110	.740	-12.92	37.35
	18	7.314	5.923	.876	-11.04	25.67
15	12	23.800(*)	5.540	.002	6.63	40.97
	13	8.400	4.331	.467	-5.02	21.82
	14	1.686	4.985	1.000	-13.76	17.13
	16	16.400(*)	4.985	.031	.95	31.85
	17	13.900	7.835	.572	-10.38	38.18
	18	9.000	5.540	.668	-8.17	26.17
16	12	7.400	5.923	.870	-10.96	25.76
	13	-8.000	4.811	.643	-22.91	6.91
	14	-14.714	5.407	.118	-31.47	2.04
	15	-16.400(*)	4.985	.031	-31.85	-.95
	17	-2.500	8.110	1.000	-27.63	22.63
	18	-7.400	5.923	.870	-25.76	10.96
17	12	9.900	8.463	.901	-16.33	36.13
	13	-5.500	7.725	.991	-29.44	18.44
	14	-12.214	8.110	.740	-37.35	12.92
	15	-13.900	7.835	.572	-38.18	10.38
	16	2.500	8.110	1.000	-22.63	27.63
	18	-4.900	8.463	.997	-31.13	21.33
18	12	14.800	6.397	.262	-5.03	34.63
	13	-.600	5.384	1.000	-17.29	16.09
	14	-7.314	5.923	.876	-25.67	11.04
	15	-9.000	5.540	.668	-26.17	8.17
	16	7.400	5.923	.870	-10.96	25.76
	17	4.900	8.463	.997	-21.33	31.13

* The mean difference is significant at the .05 level.

Table B.2.5 ANOVA (IQ, Education with age as a covariate)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power(a)
Corrected Model	2966.587(b)	7	423.798	4.302	.001	30.117	.976
Intercept	10238.053	1	10238.053	103.938	.000	103.938	1.000
Age	254.649	1	254.649	2.585	.116	2.585	.348
Education	2073.988	6	345.665	3.509	.007	21.055	.911
Error	3940.079	40	98.502				
Total	553040.000	48					
Corrected Total	6906.667	47					

a Computed using alpha = .05
b R Squared = .430 (Adjusted R Squared = .330)

Table B.2.6 ANOVA (IQ, Gender)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	207.002	1	207.002	1.421	.239
Within Groups	6699.664	46	145.645		
Total	6906.667	47			

Table B.2.7 ANOVA (IQ, Occupation)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2409.778	3	803.259	7.860	.000
Within Groups	4496.889	44	102.202		
Total	6906.667	47			

Table B.2.8 Tukey Post Hoc Comparisons for IQ, Occupation ANOVA

(I) Occupation	(J) Occupation	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	18.531(*)	4.323	.001	6.99	30.07
	3	15.460	7.429	.175	-4.38	35.30
	4	10.103(*)	3.375	.023	1.09	19.11
2	1	-18.531(*)	4.323	.001	-30.07	-6.99
	3	-3.071	8.106	.981	-24.71	18.57
	4	-8.429	4.680	.287	-20.92	4.07
3	1	-15.460	7.429	.175	-35.30	4.38
	2	3.071	8.106	.981	-18.57	24.71
	4	-5.357	7.642	.896	-25.76	15.05
4	1	-10.103(*)	3.375	.023	-19.11	-1.09
	2	8.429	4.680	.287	-4.07	20.92
	3	5.357	7.642	.896	-15.05	25.76

* The mean difference is significant at the .05 level.

Table B.2.9 ANOVA (IQ, Ethnicity)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3408.133	3	1136.044	14.288	.000
Within Groups	3498.534	44	79.512		
Total	6906.667	47			

Table B.2.10 Tukey Post Hoc Comparisons for IQ, Ethnicity ANOVA

(I) Ethnicity	(J) Ethnicity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	17.519(*)	2.959	.000	9.62	25.42
	4	23.117(*)	5.018	.000	9.72	36.51
	6	15.033(*)	4.307	.006	3.53	26.53
2	1	-17.519(*)	2.959	.000	-25.42	-9.62
	4	5.598	4.831	.656	-7.30	18.50
	6	-2.486	4.088	.929	-13.40	8.43
4	1	-23.117(*)	5.018	.000	-36.51	-9.72
	2	-5.598	4.831	.656	-18.50	7.30
	6	-8.083	5.756	.503	-23.45	7.28
6	1	-15.033(*)	4.307	.006	-26.53	-3.53
	2	2.486	4.088	.929	-8.43	13.40
	4	8.083	5.756	.503	-7.28	23.45

* The mean difference is significant at the .05 level.

Table B.2.11 Tests of Between-Subjects Effects for IQ, Ethnicity ANOVA covaried for Education

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3769.574(a)	5	753.915	10.094	.000
Intercept	2943.772	1	2943.772	39.412	.000
Education	361.358	1	361.358	4.838	.033
Ethnicity	3715.771	4	928.943	12.437	.000
Error	3137.093	42	74.693		
Total	553040.000	48			
Corrected Total	6906.667	47			

a R Squared = .546 (Adjusted R Squared = .492)

Table B.2.12 Tests of Between-Subjects Effects for IQ, Ethnicity ANOVA covaried for Education and Age

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4112.646(a)	6	685.441	10.058	.000
Intercept	3177.608	1	3177.608	46.629	.000
Education	553.991	1	553.991	8.129	.007
Age	343.073	1	343.073	5.034	.030
Ethnicity	2750.167	4	687.542	10.089	.000
Error	2794.020	41	68.147		
Total	553040.000	48			
Corrected Total	6906.667	47			

a R Squared = .595 (Adjusted R Squared = .536)

B.2.13 MANOVA for paired differences based on IQ strata

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.520	4.180(a)	7.000	27.000	.003
	Wilks' Lambda	.480	4.180(a)	7.000	27.000	.003
	Hotelling's Trace	1.084	4.180(a)	7.000	27.000	.003
	Roy's Largest Root	1.084	4.180(a)	7.000	27.000	.003
IQstrata	Pillai's Trace	.592	1.018	21.000	87.000	.451
	Wilks' Lambda	.507	.993	21.000	78.080	.482
	Hotelling's Trace	.788	.963	21.000	77.000	.516
	Roy's Largest Root	.458	1.899(b)	7.000	29.000	.106

a Exact statistic
b The statistic is an upper bound on F that yields a lower bound on the significance level.
c Design: Intercept+IQstrata

B.2.13 Tests of Between-Subjects Effects for paired differences based on IQ strata

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	WOMBAT	678.133(a)	3	226.044	.338	.798
	LNS	24.274(b)	3	8.091	2.152	.112
	LS	13.543(c)	3	4.514	.561	.645
	NS	26.045(d)	3	8.682	1.068	.376
	MTT	3.826(e)	3	1.275	.319	.811
	IGT	407.133(f)	3	135.711	.805	.500
	PVT	69017.565(g)	3	23005.855	2.464	.080
Intercept	WOMBAT	949.650	1	949.650	1.418	.242
	LNS	1.936	1	1.936	.515	.478
	LS	1.156	1	1.156	.144	.707
	NS	9.409	1	9.409	1.158	.290
	MTT	35.344	1	35.344	8.846	.005
	IGT	257.049	1	257.049	1.525	.226
	PVT	106122.138	1	106122.138	11.368	.002
IQstrata	WOMBAT	678.133	3	226.044	.338	.798
	LNS	24.274	3	8.091	2.152	.112
	LS	13.543	3	4.514	.561	.645
	NS	26.045	3	8.682	1.068	.376
	MTT	3.826	3	1.275	.319	.811
	IGT	407.133	3	135.711	.805	.500
	PVT	69017.565	3	23005.855	2.464	.080
Error	WOMBAT	22099.396	33	669.679		
	LNS	124.050	33	3.759		
	LS	265.700	33	8.052		
	NS	268.225	33	8.128		
	MTT	131.850	33	3.995		
	IGT	5560.975	33	168.514		
	PVT	308072.667	33	9335.535		
Total	WOMBAT	23215.510	37			
	LNS	149.000	37			
	LS	289.000	37			
	NS	317.000	37			
	MTT	188.000	37			
	IGT	6233.000	37			
	PVT	618544.884	37			
Corrected Total	WOMBAT	22777.529	36			
	LNS	148.324	36			
	LS	279.243	36			
	NS	294.270	36			
	MTT	135.676	36			
	IGT	5968.108	36			
	PVT	377090.232	36			

a R Squared = .030 (Adjusted R Squared = -.058)
b R Squared = .164 (Adjusted R Squared = .088)

c R Squared = .048 (Adjusted R Squared = -.038)
d R Squared = .089 (Adjusted R Squared = .006)
e R Squared = .028 (Adjusted R Squared = -.060)
f R Squared = .068 (Adjusted R Squared = -.016)
g R Squared = .183 (Adjusted R Squared = .109)

Table B.3 Comparison of Study Sample and Active Duty Military Demographics

Demographic	Percentage of Study Sample	Percentage of Active Duty		Percentage of Reserve/National Guard	
		Officer	Enlisted	Officer	Enlisted
Age					
<25	85.4	13.7	53.8	2.1	35.1
26-30	12.5	22.0	17.4	7.1	14.4
31-35	2.1	21.8	12.3	18.3	14.8
36-40	0.0	19.2	10.9	25.9	14.6
>41	0.0	23.3	5.5	46.7	21.2
Ethnicity					
White	31.3	82.0	63.5	83.1	69.1
African American	47.9	9.1	21.0	9.1	17.5
Hispanic American	8.3	4.5	9.9	4.1	9.0
Native American/Alaska Native	0.0	0.6	1.3	0.8	1.0
Asian American	12.5	3.7	4.1	2.8	3.2
Multi-racial	0.0	0.1	0.2	0.1	0.2
Gender					
Male	54.2	84.7	85.0	82.3	82.7
Female	45.8	15.3	15.0	17.7	17.3
Education					
No High School Diploma or GED	0.0	0.3	0.9	0.1	4.7
Less than Bachelor's Degree	85.4	8.5	94.0	11.7	85.3
Bachelor's Degree	4.2	52.4	3.3	51.8	7.0
Advanced Degree	10.4	33.7	0.4	31.4	0.9
Unknown	0.0	5.1	1.4	5.0	2.0

APPENDIX C. SUBJECTIVE SLEEPINESS

C.1 Karolinska vs. Stanford Scales Comparisons

Table C.1.1 Baseline Karolinska vs. Stanford T-Test Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Karolinska5	1.00	22	2.9545	1.25270	.26708
	2.00	26	2.6538	1.16421	.22832
Stanford5	1.00	22	1.9091	.68376	.14578
	2.00	26	2.0000	.89443	.17541

Table C.1.2 Baseline Karolinska vs. Stanford T-Test Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Karolinska5	Equal variances assumed	.077	.782	.861	46	.394	.30070	.34919	-.40218	1.00358
	Equal variances not assumed			.856	43.427	.397	.30070	.35137	-.40770	1.00910
Stanford5	Equal variances assumed	1.771	.190	-.390	46	.698	-.09091	.23323	-.56038	.37856
	Equal variances not assumed			-.399	45.577	.692	-.09091	.22808	-.55013	.36831

Table C.1.3 Sleep Deprived Karolinska vs. Stanford T-Test Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Karolinska14	1.00	22	6.0455	2.19257	.46746
	2.00	26	5.0385	1.92833	.37818
Stanford14	1.00	22	4.2273	1.50971	.32187
	2.00	26	3.5000	1.02956	.20191

Table C.1.4 Sleep Deprived Karolinska vs. Stanford T-Test Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Karolinska14	Equal variances assumed	.855	.360	1.693	46	.097	1.00699	.59477	-.19022	2.20421
	Equal variances not assumed			1.675	42.273	.101	1.00699	.60128	-.20620	2.22019
Stanford14	Equal variances assumed	4.591	.037	1.975	46	.054	.72727	.36832	-.01411	1.46866
	Equal variances not assumed			1.914	36.086	.064	.72727	.37996	-.04326	1.49781

C.2 Sleepiness and Demographics

Table C.2.1 Sleepiness and Occupation Descriptives

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Karolinska5	1.00	25	2.9200	1.22202	.24440	2.4156	3.4244	1.00	5.00
	2.00	7	2.5714	1.27242	.48093	1.3946	3.7482	1.00	5.00
	3.00	2	2.0000	.00000	.00000	2.0000	2.0000	2.00	2.00
	4.00	14	2.7857	1.25137	.33444	2.0632	3.5082	1.00	6.00
	Total	48	2.7917	1.20210	.17351	2.4426	3.1407	1.00	6.00
Stanford5	1.00	25	2.0000	.81650	.16330	1.6630	2.3370	1.00	4.00
	2.00	7	1.4286	.78680	.29738	.7009	2.1562	1.00	3.00
	3.00	2	1.5000	.70711	.50000	-4.8531	7.8531	1.00	2.00
	4.00	14	2.2143	.69929	.18689	1.8105	2.6180	1.00	3.00
	Total	48	1.9583	.79783	.11516	1.7267	2.1900	1.00	4.00
Karolinska14	1.00	25	6.0800	1.97737	.39547	5.2638	6.8962	3.00	9.00
	2.00	7	4.2857	2.21467	.83707	2.2375	6.3339	2.00	8.00
	3.00	2	4.5000	2.12132	1.50000	-14.5593	23.5593	3.00	6.00
	4.00	14	5.2143	2.08211	.55647	4.0121	6.4165	2.00	8.00
	Total	48	5.5000	2.09356	.30218	4.8921	6.1079	2.00	9.00
Stanford14	1.00	25	4.2800	1.13725	.22745	3.8106	4.7494	3.00	6.00
	2.00	7	2.7143	1.25357	.47380	1.5549	3.8736	1.00	4.00
	3.00	2	3.0000	1.41421	1.00000	-9.7062	15.7062	2.00	4.00
	4.00	14	3.7143	1.32599	.35438	2.9487	4.4799	1.00	6.00
	Total	48	3.8333	1.31008	.18909	3.4529	4.2137	1.00	6.00

Table C.2.2 Sleepiness and Occupation ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Karolinska5	Between Groups	2.005	3	.668	.446	.721
	Within Groups	65.911	44	1.498		
	Total	67.917	47			
Stanford5	Between Groups	3.345	3	1.115	1.846	.153
	Within Groups	26.571	44	.604		
	Total	29.917	47			
Karolinska14	Between Groups	21.874	3	7.291	1.742	.172
	Within Groups	184.126	44	4.185		
	Total	206.000	47			
Stanford14	Between Groups	15.341	3	5.114	3.444	.025
	Within Groups	65.326	44	1.485		
	Total	80.667	47			

Table C.2.3 Sleepiness and Occupation Tukey Multiple Comparisons

Dependent Variable	(I) Occupation	(J) Occupation	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Karolinska5	1.00	2.00	.34857	.52337	.909	-1.0488	1.7460
		3.00	.92000	.89940	.737	-1.4814	3.3214
		4.00	.13429	.40856	.988	-.9566	1.2251
	2.00	1.00	-.34857	.52337	.909	-1.7460	1.0488
		3.00	.57143	.98132	.937	-2.0487	3.1916
		4.00	-.21429	.56657	.981	-1.7270	1.2984
	3.00	1.00	-.92000	.89940	.737	-3.3214	1.4814
		2.00	-.57143	.98132	.937	-3.1916	2.0487
		4.00	-.78571	.92520	.831	-3.2560	1.6846
	4.00	1.00	-.13429	.40856	.988	-1.2251	.9566
		2.00	.21429	.56657	.981	-1.2984	1.7270
		3.00	.78571	.92520	.831	-1.6846	3.2560
Stanford5	1.00	2.00	.57143	.33231	.326	-.3158	1.4587
		3.00	.50000	.57106	.817	-1.0247	2.0247
		4.00	-.21429	.25941	.842	-.9069	.4783
	2.00	1.00	-.57143	.33231	.326	-1.4587	.3158
		3.00	-.07143	.62307	.999	-1.7350	1.5922
		4.00	-.78571	.35973	.144	-1.7462	.1748
	3.00	1.00	-.50000	.57106	.817	-2.0247	1.0247
		2.00	.07143	.62307	.999	-1.5922	1.7350
		4.00	-.71429	.58744	.620	-2.2827	.8542
	4.00	1.00	.21429	.25941	.842	-.4783	.9069
		2.00	.78571	.35973	.144	-.1748	1.7462
		3.00	.71429	.58744	.620	-.8542	2.2827

Karolinska14	1.00	2.00	1.79429	.87476	.185	-.5413	4.1299
		3.00	1.58000	1.50324	.721	-2.4337	5.5937
		4.00	.86571	.68286	.588	-.9575	2.6889
	2.00	1.00	-1.79429	.87476	.185	-4.1299	.5413
		3.00	-.21429	1.64017	.999	-4.5935	4.1650
		4.00	-.92857	.94695	.761	-3.4569	1.5998
	3.00	1.00	-1.58000	1.50324	.721	-5.5937	2.4337
		2.00	.21429	1.64017	.999	-4.1650	4.5935
		4.00	-.71429	1.54636	.967	-4.8431	3.4145
	4.00	1.00	-.86571	.68286	.588	-2.6889	.9575
		2.00	.92857	.94695	.761	-1.5998	3.4569
		3.00	.71429	1.54636	.967	-3.4145	4.8431
Stanford14	1.00	2.00	1.56571(*)	.52104	.022	.1745	2.9569
		3.00	1.28000	.89539	.488	-1.1107	3.6707
		4.00	.56571	.40674	.512	-.5203	1.6517
	2.00	1.00	-1.56571(*)	.52104	.022	-2.9569	-.1745
		3.00	-.28571	.97695	.991	-2.8942	2.3228
		4.00	-1.00000	.56404	.300	-2.5060	.5060
	3.00	1.00	-1.28000	.89539	.488	-3.6707	1.1107
		2.00	.28571	.97695	.991	-2.3228	2.8942
		4.00	-.71429	.92108	.865	-3.1736	1.7450
	4.00	1.00	-.56571	.40674	.512	-1.6517	.5203
		2.00	1.00000	.56404	.300	-.5060	2.5060
		3.00	.71429	.92108	.865	-1.7450	3.1736

* The mean difference is significant at the .05 level.

Table C.2.4 Sleepiness and Morningness-Eveningness Descriptives

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Karolinska5	1.00	14	2.7143	1.06904	.28571	2.0970	3.3315	1.00	5.00
	2.00	27	2.5556	1.12090	.21572	2.1121	2.9990	1.00	5.00
	3.00	7	3.8571	1.34519	.50843	2.6131	5.1012	2.00	6.00
	Total	48	2.7917	1.20210	.17351	2.4426	3.1407	1.00	6.00
Stanford5	1.00	14	2.0000	.96077	.25678	1.4453	2.5547	1.00	4.00
	2.00	27	1.8148	.73574	.14159	1.5238	2.1059	1.00	3.00
	3.00	7	2.4286	.53452	.20203	1.9342	2.9229	2.00	3.00
	Total	48	1.9583	.79783	.11516	1.7267	2.1900	1.00	4.00
Karolinska14	1.00	14	4.6429	1.54955	.41413	3.7482	5.5375	2.00	8.00
	2.00	27	5.6296	2.22137	.42750	4.7509	6.5084	2.00	9.00
	3.00	7	6.7143	2.05866	.77810	4.8103	8.6182	3.00	9.00
	Total	48	5.5000	2.09356	.30218	4.8921	6.1079	2.00	9.00
Stanford14	1.00	14	3.2857	1.13873	.30434	2.6282	3.9432	1.00	5.00
	2.00	27	3.9259	1.23805	.23826	3.4362	4.4157	1.00	6.00
	3.00	7	4.5714	1.61835	.61168	3.0747	6.0682	2.00	6.00
	Total	48	3.8333	1.31008	.18909	3.4529	4.2137	1.00	6.00

Table C.2.5 Sleepiness and Morningness-Eveningness ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Karolinska5	Between Groups	9.536	2	4.768	3.675	.033
	Within Groups	58.381	45	1.297		
	Total	67.917	47			
Stanford5	Between Groups	2.128	2	1.064	1.723	.190
	Within Groups	27.788	45	.618		
	Total	29.917	47			
Karolinska14	Between Groups	21.061	2	10.530	2.562	.088
	Within Groups	184.939	45	4.110		
	Total	206.000	47			
Stanford14	Between Groups	8.243	2	4.122	2.561	.088
	Within Groups	72.423	45	1.609		
	Total	80.667	47			

Table C.2.6 Sleepiness and Morningness-Eveningness Tukey Multiple Comparisons

Dependent Variable	(I) MEQtype	(J) MEQtype	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Karolinska5	1.00	2.00	.15873	.37512	.906	-.7504	1.0679
		3.00	-1.14286	.52726	.088	-2.4207	.1350
	2.00	1.00	-.15873	.37512	.906	-1.0679	.7504
		3.00	-1.30159(*)	.48310	.026	-2.4724	-.1307
	3.00	1.00	1.14286	.52726	.088	-.1350	2.4207
		2.00	1.30159(*)	.48310	.026	.1307	2.4724
Stanford5	1.00	2.00	.18519	.25880	.756	-.4421	.8124
		3.00	-.42857	.36377	.472	-1.3102	.4531
	2.00	1.00	-.18519	.25880	.756	-.8124	.4421
		3.00	-.61376	.33330	.168	-1.4215	.1940
	3.00	1.00	.42857	.36377	.472	-.4531	1.3102
		2.00	.61376	.33330	.168	-.1940	1.4215
Karolinska14	1.00	2.00	-.98677	.66766	.311	-2.6049	.6314
		3.00	-2.07143	.93844	.081	-4.3458	.2030
	2.00	1.00	.98677	.66766	.311	-.6314	2.6049
		3.00	-1.08466	.85984	.424	-3.1686	.9993
	3.00	1.00	2.07143	.93844	.081	-.2030	4.3458
		2.00	1.08466	.85984	.424	-.9993	3.1686
Stanford14	1.00	2.00	-.64021	.41781	.286	-1.6528	.3724
		3.00	-1.28571	.58726	.084	-2.7090	.1376
	2.00	1.00	.64021	.41781	.286	-.3724	1.6528
		3.00	-.64550	.53807	.460	-1.9496	.6586
	3.00	1.00	1.28571	.58726	.084	-.1376	2.7090
		2.00	.64550	.53807	.460	-.6586	1.9496

* The mean difference is significant at the .05 level.

C.3 Polyomial Regression Analysis (mean RT vs. Karolinska Scale)

```
The regression equation is
mean RT = 385.0 - 88.52 Mean Karolinska + 13.57 Mean Karolinska**2


S = 11.1347   R-Sq = 88.8%   R-Sq(adj) = 86.9%


Analysis of Variance

Source       DF       SS       MS      F      P
Regression    2  11761.3  5880.63  47.43  0.000
Error        12   1487.8   123.98
Total        14  13249.0


Sequential Analysis of Variance

Source     DF       SS      F      P

Linear      1  10157.7  42.72  0.000
Quadratic   1   1603.6  12.93  0.004
```

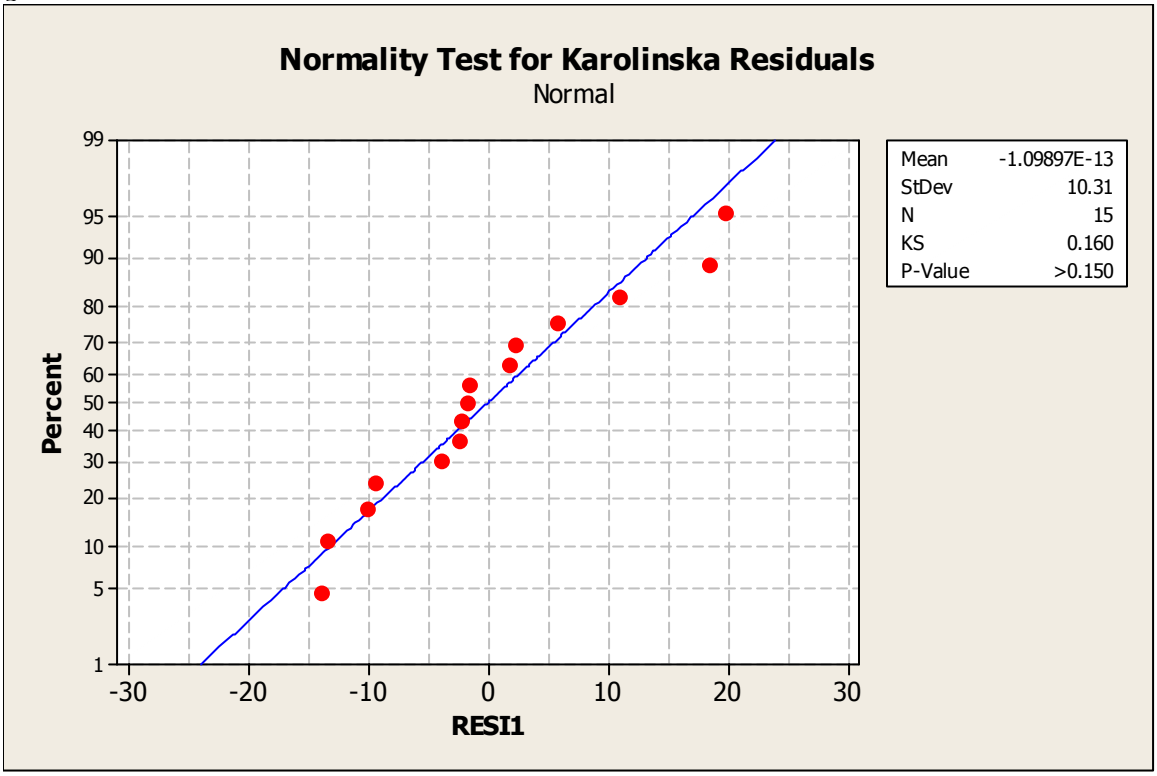


Figure C.3.1 Normality Test for Karolinska Residuals

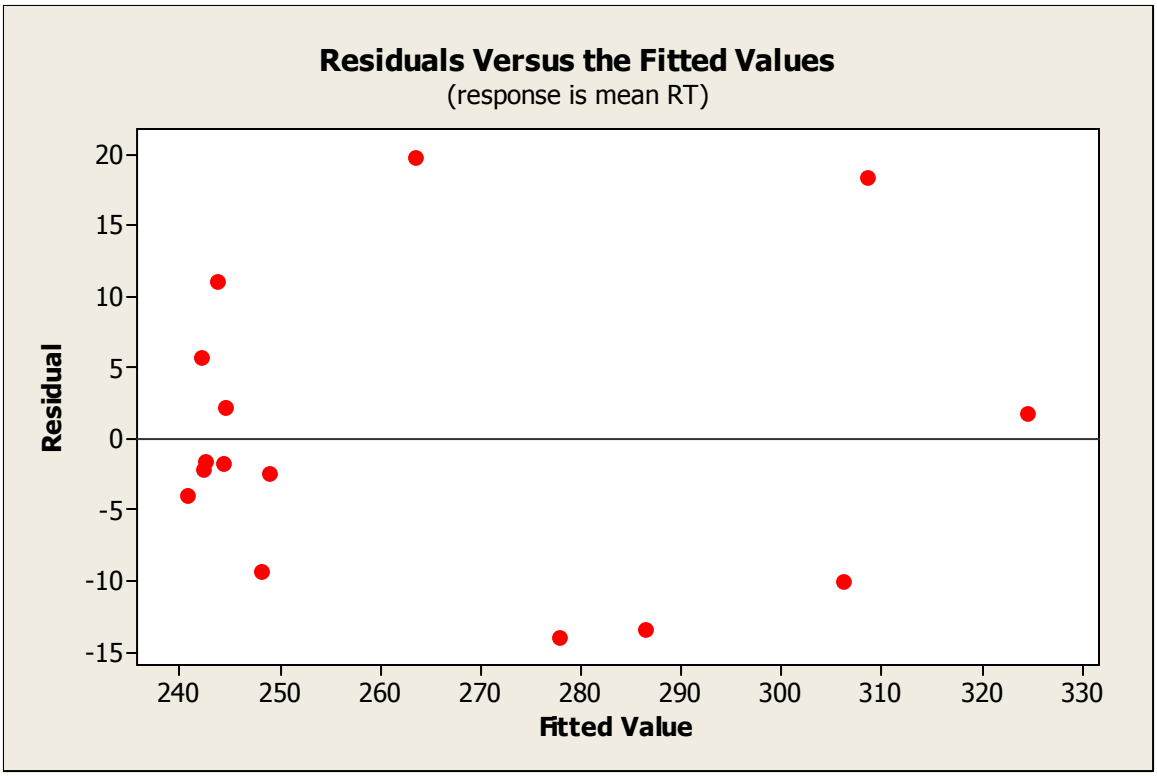


Figure C.3.2 Residual vs. Fitted Values (Karolinska Scale)

C.4 Polyomial Regression Analysis (mean RT vs. Stanford Scale)

```
The regression equation is
mean RT = 303.8 - 59.88 Mean Stanford + 14.90 Mean Stanford**2


S = 15.2303   R-Sq = 79.0%   R-Sq(adj) = 75.5%


Analysis of Variance

Source       DF       SS       MS      F      P
Regression    2  10465.5  5232.74  22.56  0.000
Error        12   2783.6   231.96
Total        14  13249.0


Sequential Analysis of Variance

Source     DF       SS      F      P
Linear      1  9639.12  34.71  0.000
Quadratic   1   826.36   3.56  0.084
```

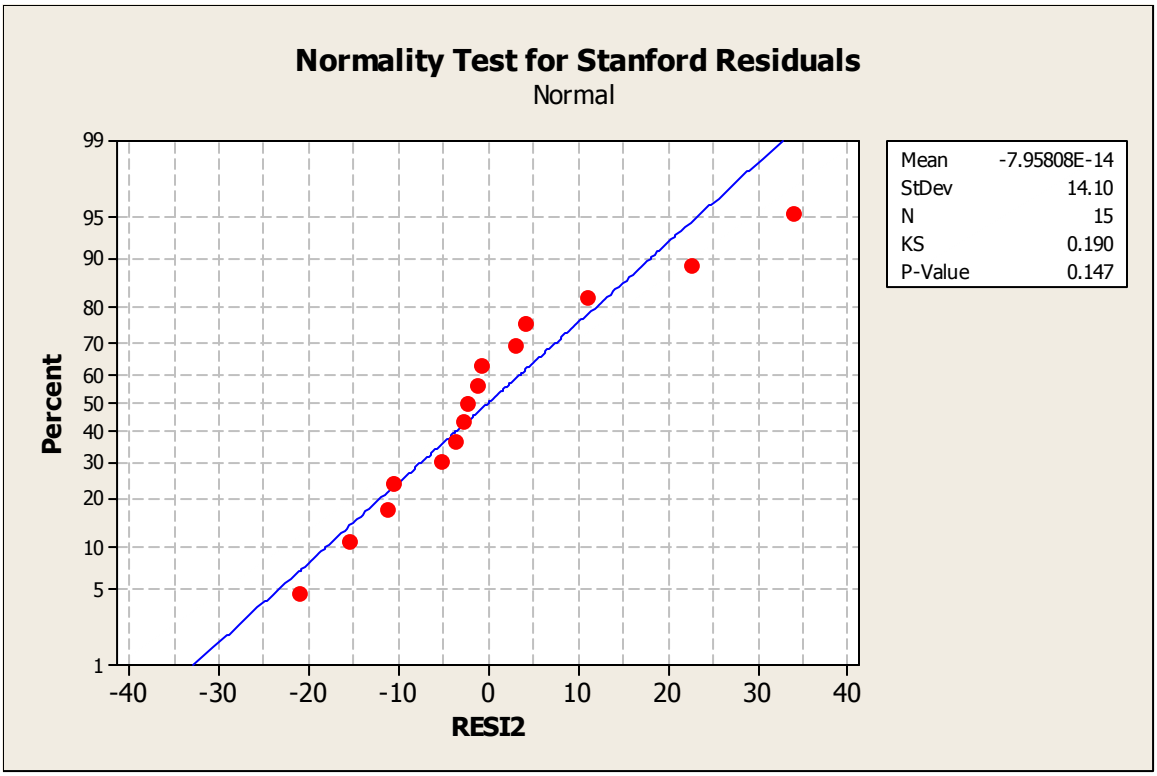


Figure C.4.1 Normality Test for Stanford Residuals

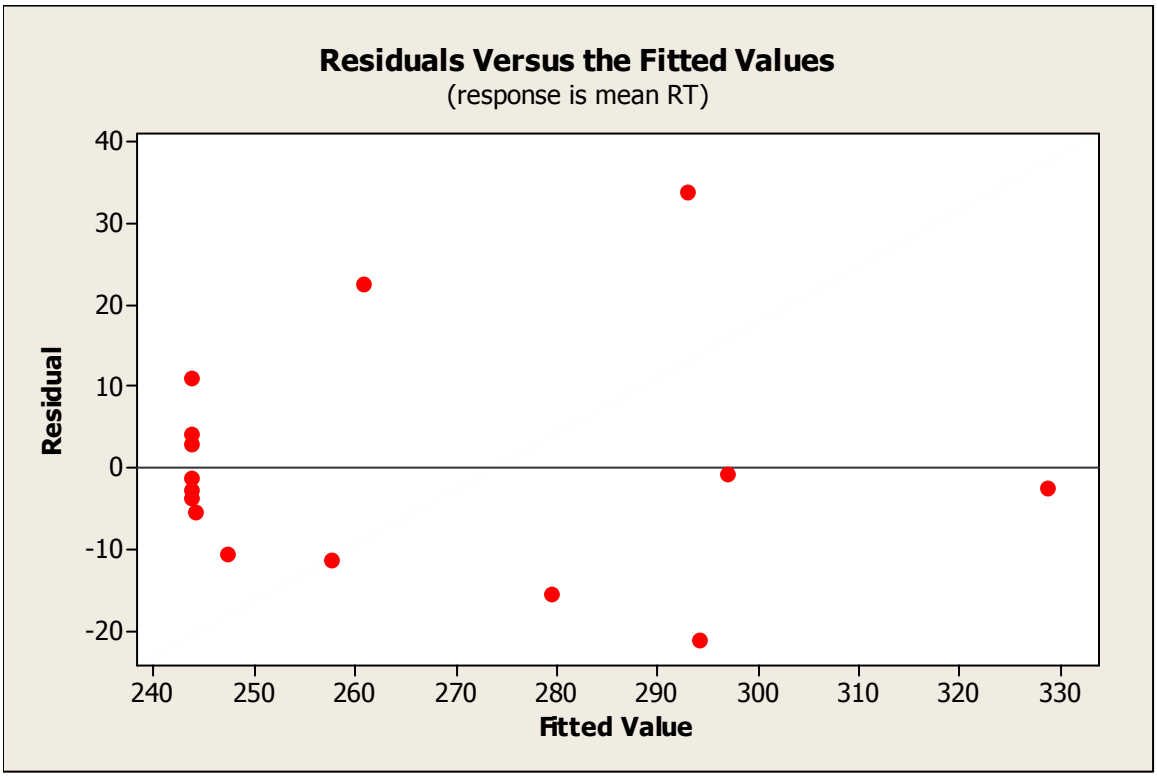


Figure C.4.2 Residual vs. Fitted Values (Stanford Scale)

APPENDIX D. COGNITIVE TEST DATA CHARACTERISTICS

D.1 WOMBAT Data (Normality and Equality of Variance)

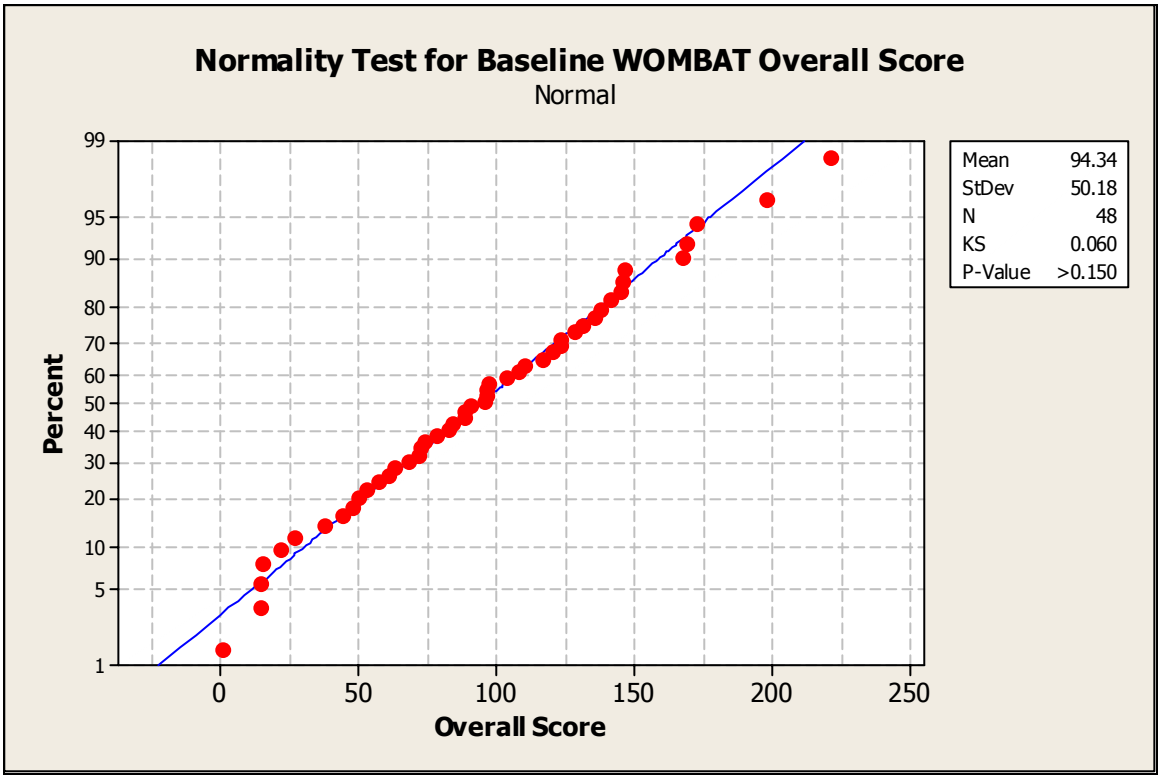


Figure D.1.1. Normality Test for Baseline WOMBAT Overall Score

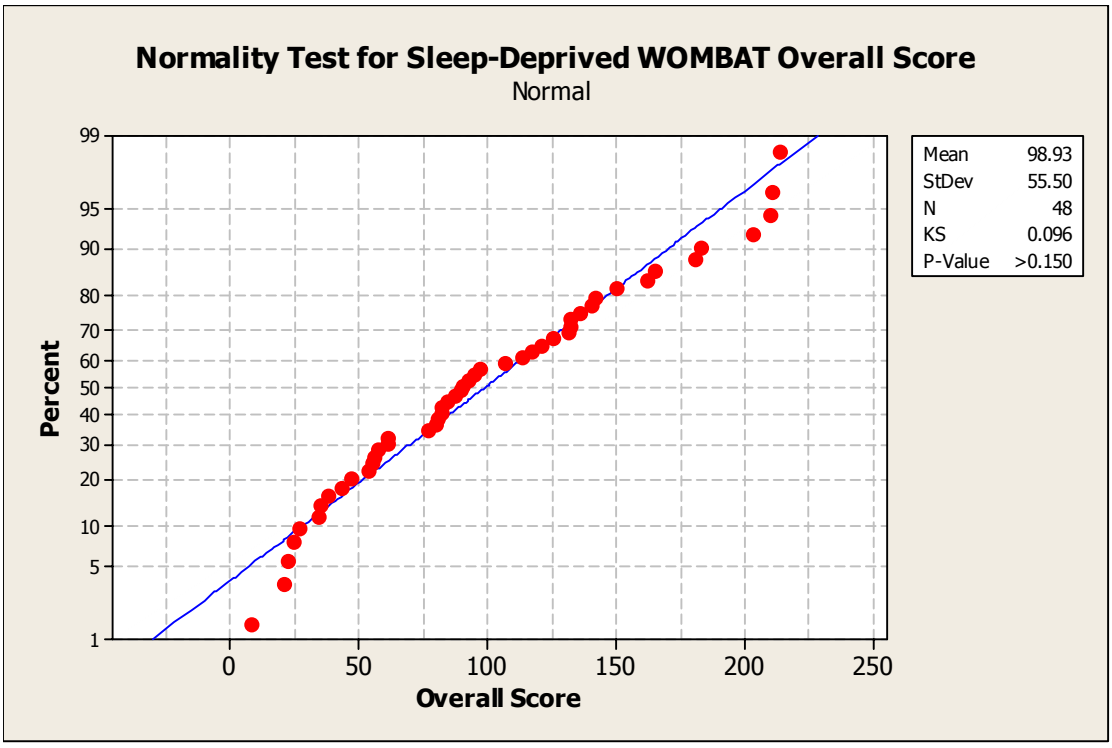


Figure D.1.2. Normality Test for Sleep-Deprived WOMBAT Overall Score

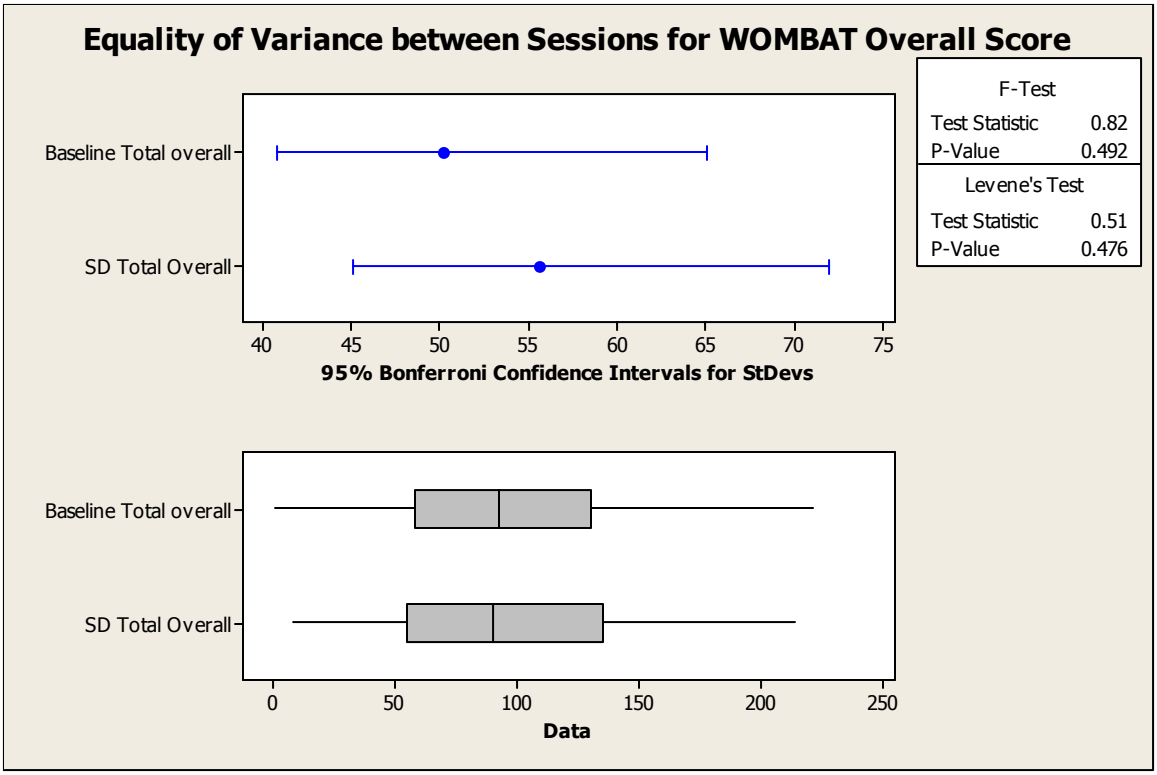


Figure D.1.3. Test for Equality of Variance for WOMBAT Overall Score

D.2 WOMBAT Improvement vs. Degradation

Table D.2.1 T-test for Baseline Scores (Improved vs. Degraded)

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Improved*	*Degraded*
Mean	103.2272727	86.81538462
Variance	2409.230649	2581.633354
Observations	22	26
Pooled Variance	2502.927771	
Hypothesized Mean Difference	0	
df	46	
t Stat	1.132431092	
P(T<=t) one-tail	0.131660873	
t Critical one-tail	1.678660414	
P(T<=t) two-tail	0.263321746	
t Critical two-tail	2.012895567	

D.2.1 Demographics and Improvement vs. Degradation

Table D.2.1.1 ANOVA for Degraded vs. Improved and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
age	Between Groups	.597	1	.597	.046	.830
	Within Groups	592.070	46	12.871		
	Total	592.667	47			
IQ	Between Groups	62.695	1	62.695	.421	.519
	Within Groups	6843.972	46	148.782		
	Total	6906.667	47			
education	Between Groups	5.149	1	5.149	1.617	.210
	Within Groups	146.517	46	3.185		
	Total	151.667	47			
ME	Between Groups	56.364	1	56.364	.612	.438
	Within Groups	4233.552	46	92.034		
	Total	4289.917	47			

Table D.2.1.2 Chi-Square Tests for Degraded vs. Improved by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
1	18	3	5	26
	18.958	2.7083	4.3333	
	-0.958	0.2917	0.6667	
	37.50	6.25	10.42	54.17
	69.23	11.54	19.23	
	51.43	60.00	62.50	
2	17	2	3	22
	16.042	2.2917	3.6667	
	0.9583	-0.292	-0.667	
	35.42	4.17	6.25	45.83
	77.27	9.09	13.64	
	48.57	40.00	37.50	
Total	35	5	8	48
	72.92	10.42	16.67	100.00

Table D.2.1.3 Chi-Square Tests for Degraded vs. Improved by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	2	0.3980	0.8195
Likelihood Ratio Chi-Square	2	0.4016	0.8181
Mantel-Haenszel Chi-Square	1	0.3745	0.5406
Phi Coefficient		0.0911	
Contingency Coefficient		0.0907	
Cramer's V		0.0911	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.3745	0.5406
2	Row Mean Scores Differ	1	0.3745	0.5406
3	General Association	2	0.3897	0.8230

Total Sample Size = 48

Table D.2.1.4 Chi-Square Tests for Degraded vs. Improved by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
1	12 11.917 0.0833 25.00 46.15 54.55	14 14.083 -0.083 29.17 53.85 53.85	26 54.17
2	10 10.083 -0.083 20.83 45.45 45.45	12 11.917 0.0833 25.00 54.55 46.15	22 45.83
Total	22 45.83	26 54.17	48 100.00

Table D.2.1.5 Chi-Square Tests for Degraded vs. Improved by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	1	0.0023	0.9614
Likelihood Ratio Chi-Square	1	0.0023	0.9614
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0023	0.9618
Phi Coefficient		0.0070	
Contingency Coefficient		0.0070	
Cramer's V		0.0070	

Fisher's Exact Test

Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.6323
Right-sided Pr >= F	0.5958
Table Probability (P)	0.2280
Two-sided Pr <= P	1.0000

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0023	0.9618
2	Row Mean Scores Differ	1	0.0023	0.9618
3	General Association	1	0.0023	0.9618

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	1.0286	0.3291	3.2149
	Logit	1.0286	0.3291	3.2149
Cohort (Col1 Risk)	Mantel-Haenszel	1.0154	0.5473	1.8837
	Logit	1.0154	0.5473	1.8837
Cohort (Col2 Risk)	Mantel-Haenszel	0.9872	0.5859	1.6633
	Logit	0.9872	0.5859	1.6633

Total Sample Size = 48

Table D.2.1.6 Chi-Square Tests for Degraded vs. Improved by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
1	15	5	1	5	26
	13.542	3.7917	1.0833	7.5833	
	1.4583	1.2083	-0.083	-2.583	
	31.25	10.42	2.08	10.42	54.17
	57.69	19.23	3.85	19.23	
	60.00	71.43	50.00	35.71	
2	10	2	1	9	22
	11.458	3.2083	0.9167	6.4167	
	-1.458	-1.208	0.0833	2.5833	
	20.83	4.17	2.08	18.75	45.83
	45.45	9.09	4.55	40.91	
	40.00	28.57	50.00	64.29	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.2.1.7 Chi-Square Tests for Degraded vs. Improved by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	3	3.1169	0.3740
Likelihood Ratio Chi-Square	3	3.1603	0.3676
Mantel-Haenszel Chi-Square	1	2.1519	0.1424
Phi Coefficient		0.2548	
Contingency Coefficient		0.2469	
Cramer's V		0.2548	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.1519	0.1424
2	Row Mean Scores Differ	1	2.1519	0.1424
3	General Association	3	3.0519	0.3837

Total Sample Size = 48

Table D.2.1.8 Chi-Square Tests for Degraded vs. Improved by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	5	6	Total
1	9	11	1	1	4	26
	8.125	12.458	1.625	0.5417	3.25	
	0.875	-1.458	-0.625	0.4583	0.75	
	18.75	22.92	2.08	2.08	8.33	54.17
	34.62	42.31	3.85	3.85	15.38	
	60.00	47.83	33.33	100.00	66.67	
2	6	12	2	0	2	22
	6.875	10.542	1.375	0.4583	2.75	
	-0.875	1.4583	0.625	-0.458	-0.75	
	12.50	25.00	4.17	0.00	4.17	45.83
	27.27	54.55	9.09	0.00	9.09	
	40.00	52.17	66.67	0.00	33.33	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.2.1.9 Chi-Square Tests for Degraded vs. Improved by Ethnicity Cross Tabulation Matrix

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	4	2.3263	0.6760
Likelihood Ratio Chi-Square	4	2.7195	0.6058
Mantel-Haenszel Chi-Square	1	0.1594	0.6897
Phi Coefficient		0.2201	
Contingency Coefficient		0.2150	
Cramer's V		0.2201	

WARNING: 60% of the cells have expected counts less
than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1594	0.6897
2	Row Mean Scores Differ	1	0.1594	0.6897
3	General Association	4	2.2778	0.6848

Total Sample Size = 48

Table D.2.1.10 Chi-Square Tests for Degraded vs. Improved by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
1	9	11	6	26
	7.5833	14.625	3.7917	
	1.4167	-3.625	2.2083	
	18.75	22.92	12.50	54.17
	34.62	42.31	23.08	
	64.29	40.74	85.71	
2	5	16	1	22
	6.4167	12.375	3.2083	
	-1.417	3.625	-2.208	
	10.42	33.33	2.08	45.83
	22.73	72.73	4.55	
	35.71	59.26	14.29	
Total	14	27	7	48
	29.17	56.25	14.58	100.00

Table D.2.1.11 Chi-Square Tests for Degraded vs. Improved by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	2	5.3440	0.0691
Likelihood Ratio Chi-Square	2	5.7189	0.0573
Mantel-Haenszel Chi-Square	1	0.1237	0.7250
Phi Coefficient		0.3337	
Contingency Coefficient		0.3165	
Cramer's V		0.3337	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1237	0.7250
2	Row Mean Scores Differ	1	0.1237	0.7250
3	General Association	2	5.2327	0.0731

Total Sample Size = 48

Table D.2.1.12 MANOVA for Degraded vs. Improved by WASI Score

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.993	204.226(a)	14.000	19.000	.000
	Wilks' Lambda	.007	204.226(a)	14.000	19.000	.000
	Hotelling's Trace	150.482	204.226(a)	14.000	19.000	.000
	Roy's Largest Root	150.482	204.226(a)	14.000	19.000	.000
IQstrata	Pillai's Trace	1.560	1.624	42.000	63.000	.040
	Wilks' Lambda	.030	3.094	42.000	57.128	.000
	Hotelling's Trace	15.580	6.554	42.000	53.000	.000
	Roy's Largest Root	14.603	21.905(b)	14.000	21.000	.000

a Exact statistic
b The statistic is an upper bound on F that yields a lower bound on the significance level.
c Design: Intercept+IQstrata

Table D.2.1.13 MANOVA for Degraded vs. Improved by WASI Score Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	BASEWOMBAT	9597.419(a)	3	3199.140	1.171	.336
	SDWOMBAT	4310.472(b)	3	1436.824	.455	.715
	BASELNS	8.382(c)	3	2.794	.834	.485
	SDLNS	2.906(d)	3	.969	.254	.858
	BASELS	31.067(e)	3	10.356	1.322	.284
	SDLS	10.267(f)	3	3.422	.584	.630
	BASENS	186.494(g)	3	62.165	77.068	.000
	SDNS	80.821(h)	3	26.940	3.168	.038
	BASEMTT	236.109(i)	3	78.703	23.865	.000
	SDMTT	223.808(j)	3	74.603	13.681	.000
	BASEIGT	329.743(k)	3	109.914	.465	.709
	SDIGT	59.646(l)	3	19.882	.078	.971
	BASEPVT	30473.062(m)	3	10157.687	.443	.724
	SDPVT	2330.710(n)	3	776.903	.035	.991
Intercept	BASEWOMBAT	206133.438	1	206133.438	75.436	.000
	SDWOMBAT	204494.958	1	204494.958	64.796	.000
	BASELNS	1652.393	1	1652.393	492.992	.000
	SDLNS	1867.451	1	1867.451	489.892	.000
	BASELS	2590.522	1	2590.522	330.829	.000
	SDLS	2841.114	1	2841.114	485.214	.000
	BASENS	1606.270	1	1606.270	1991.359	.000
	SDNS	1389.656	1	1389.656	163.398	.000
	BASEMTT	3520.960	1	3520.960	1067.663	.000

	SDMTT	4645.585	1	4645.585	851.925	.000
	BASEIGT	37031.607	1	37031.607	156.639	.000
	SDIGT	44191.965	1	44191.965	173.571	.000
	BASEPVT	3160676.189	1	3160676.189	137.938	.000
	SDPVT	257563.412	1	257563.412	11.756	.002
	BASEWOMBAT	9597.419	3	3199.140	1.171	.336
	SDWOMBAT	4310.472	3	1436.824	.455	.715
	BASELNS	8.382	3	2.794	.834	.485
	SDLNS	2.906	3	.969	.254	.858
	BASELS	31.067	3	10.356	1.322	.284
	SDLS	10.267	3	3.422	.584	.630
IQstrata	BASENS	186.494	3	62.165	77.068	.000
	SDNS	80.821	3	26.940	3.168	.038
	BASEMTT	236.109	3	78.703	23.865	.000
	SDMTT	223.808	3	74.603	13.681	.000
	BASEIGT	329.743	3	109.914	.465	.709
	SDIGT	59.646	3	19.882	.078	.971
	BASEPVT	30473.062	3	10157.687	.443	.724
	SDPVT	2330.710	3	776.903	.035	.991
	BASEWOMBAT	87441.508	32	2732.547		
	SDWOMBAT	100991.411	32	3155.982		
	BASELNS	107.257	32	3.352		
	SDLNS	121.983	32	3.812		
	BASELS	250.572	32	7.830		
	SDLS	187.372	32	5.855		
Error	BASENS	25.812	32	.807		
	SDNS	272.151	32	8.505		
	BASEMTT	105.530	32	3.298		
	SDMTT	174.497	32	5.453		
	BASEIGT	7565.257	32	236.414		
	SDIGT	8147.326	32	254.604		
	BASEPVT	733237.246	32	22913.664		
	SDPVT	701068.822	32	21908.401		
	BASEWOMBAT	386429.130	36			
	SDWOMBAT	411239.930	36			
	BASELNS	2533.000	36			
	SDLNS	2692.000	36			
	BASELS	4105.000	36			
	SDLS	4315.000	36			
Total	BASENS	2313.000	36			
	SDNS	2131.000	36			
	BASEMTT	5265.000	36			
	SDMTT	6613.000	36			
	BASEIGT	58520.000	36			
	SDIGT	68641.000	36			
	BASEPVT	5231304.131	36			
	SDPVT	1075572.736	36			

Corrected Total	BASEWOMBAT	97038.927	35			
	SDWOMBAT	105301.883	35			
	BASELNS	115.639	35			
	SDLNS	124.889	35			
	BASELS	281.639	35			
	SDLS	197.639	35			
	BASENS	212.306	35			
	SDNS	352.972	35			
	BASEMTT	341.639	35			
	SDMTT	398.306	35			
	BASEIGT	7895.000	35			
	SDIGT	8206.972	35			
	BASEPVT	763710.308	35			
	SDPVT	703399.532	35			

a R Squared = .099 (Adjusted R Squared = .014)
b R Squared = .041 (Adjusted R Squared = -.049)
c R Squared = .072 (Adjusted R Squared = -.014)
d R Squared = .023 (Adjusted R Squared = -.068)
e R Squared = .110 (Adjusted R Squared = .027)
f R Squared = .052 (Adjusted R Squared = -.037)
g R Squared = .878 (Adjusted R Squared = .867)
h R Squared = .229 (Adjusted R Squared = .157)
i R Squared = .691 (Adjusted R Squared = .662)
j R Squared = .562 (Adjusted R Squared = .521)
k R Squared = .042 (Adjusted R Squared = -.048)
l R Squared = .007 (Adjusted R Squared = -.086)
m R Squared = .040 (Adjusted R Squared = -.050)
n R Squared = .003 (Adjusted R Squared = -.090)

D.3 Letter-Number Sequencing Data (Normality and Equality of Variance)

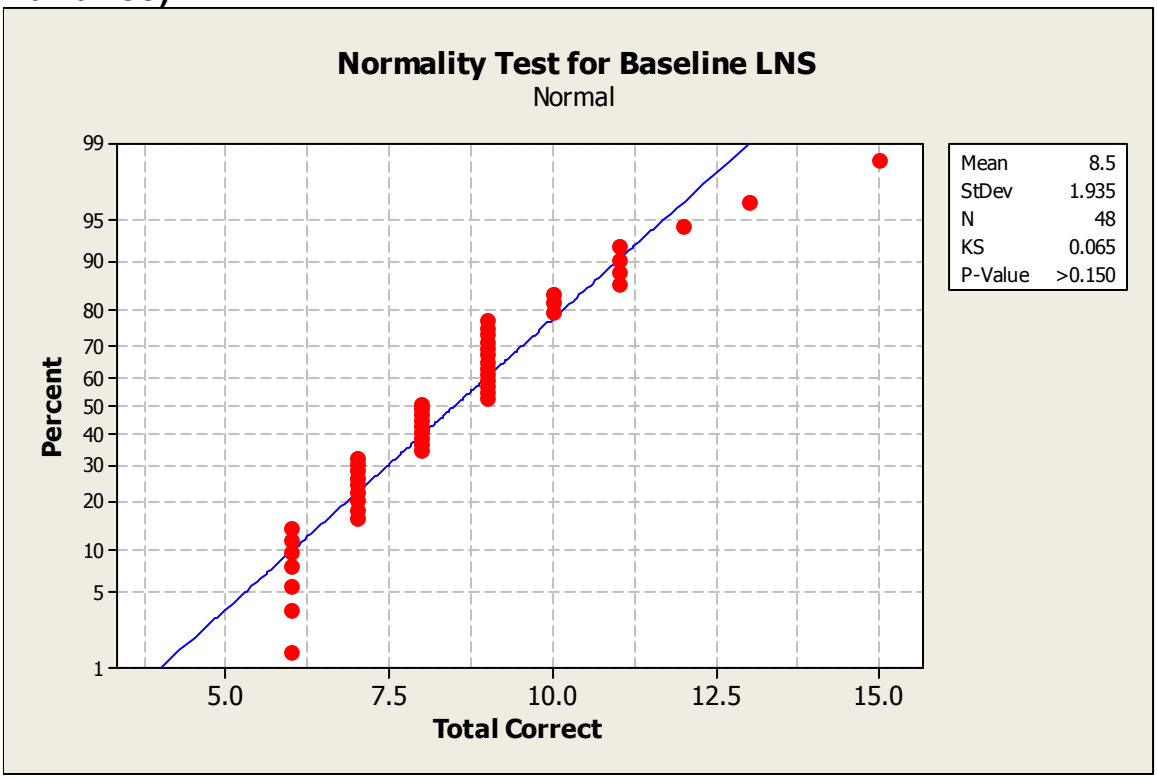


Figure D.3.1. Normality Test for Baseline Letter-Number Sequencing

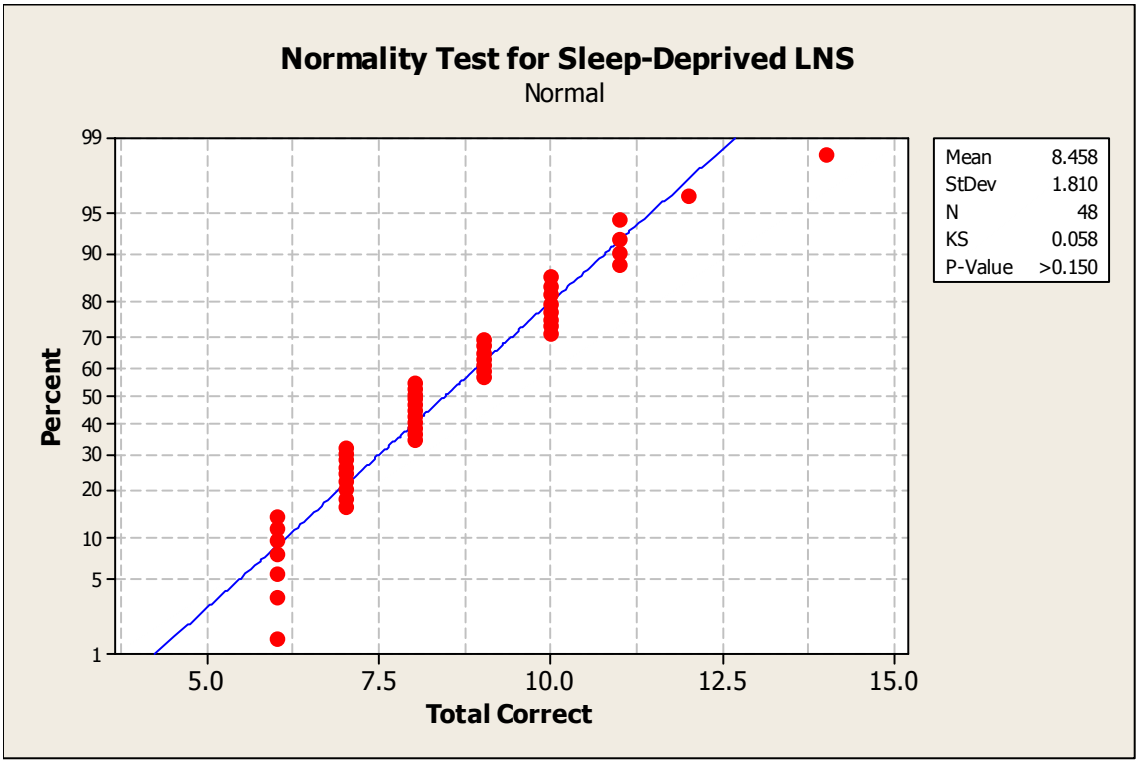


Figure D.3.2. Normality Test for Sleep-Deprived Letter-Number Sequencing

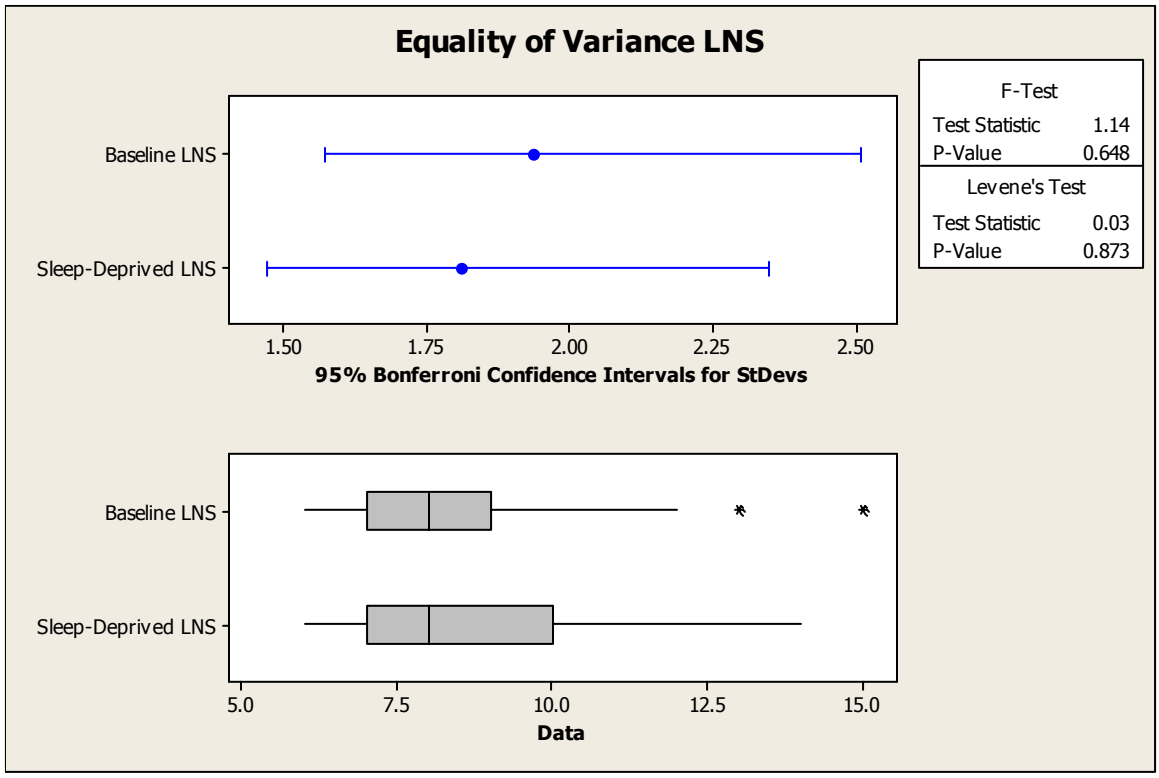


Figure D.3.3. Test for Equality of Variance for Letter Number Sequencing

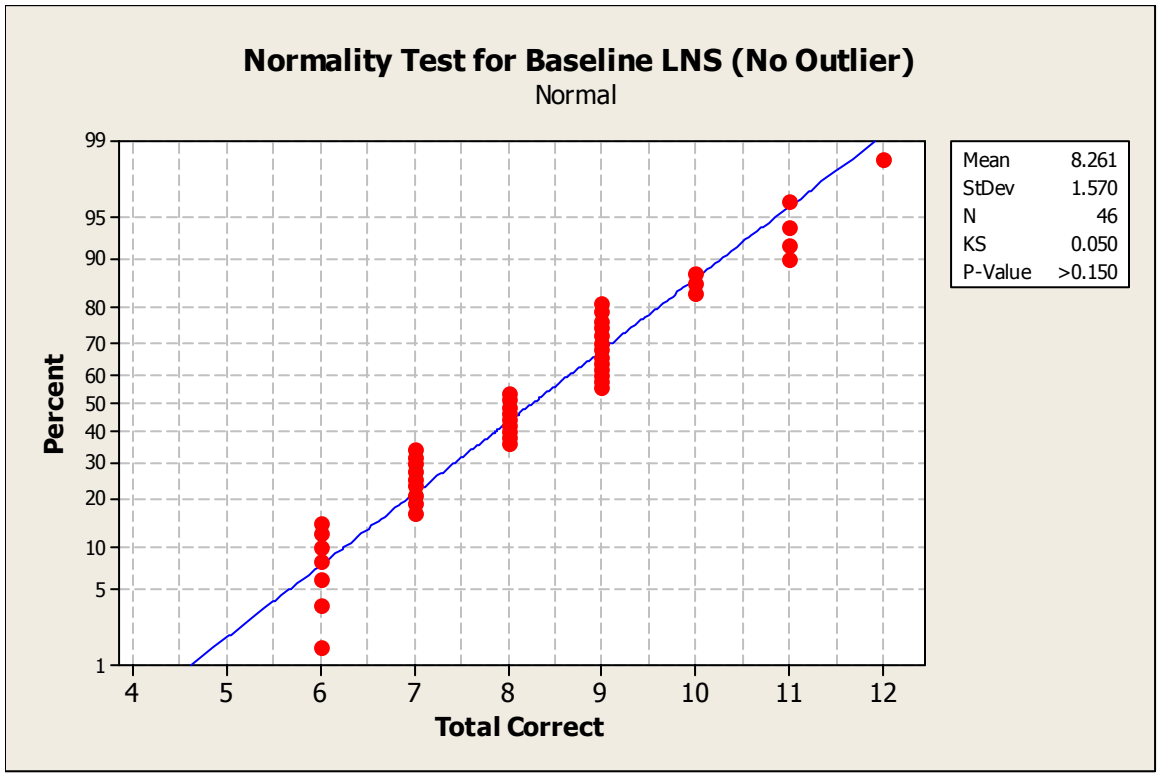


Figure D.3.4. Normality Test for Baseline Letter-Number Sequencing with No Outliers

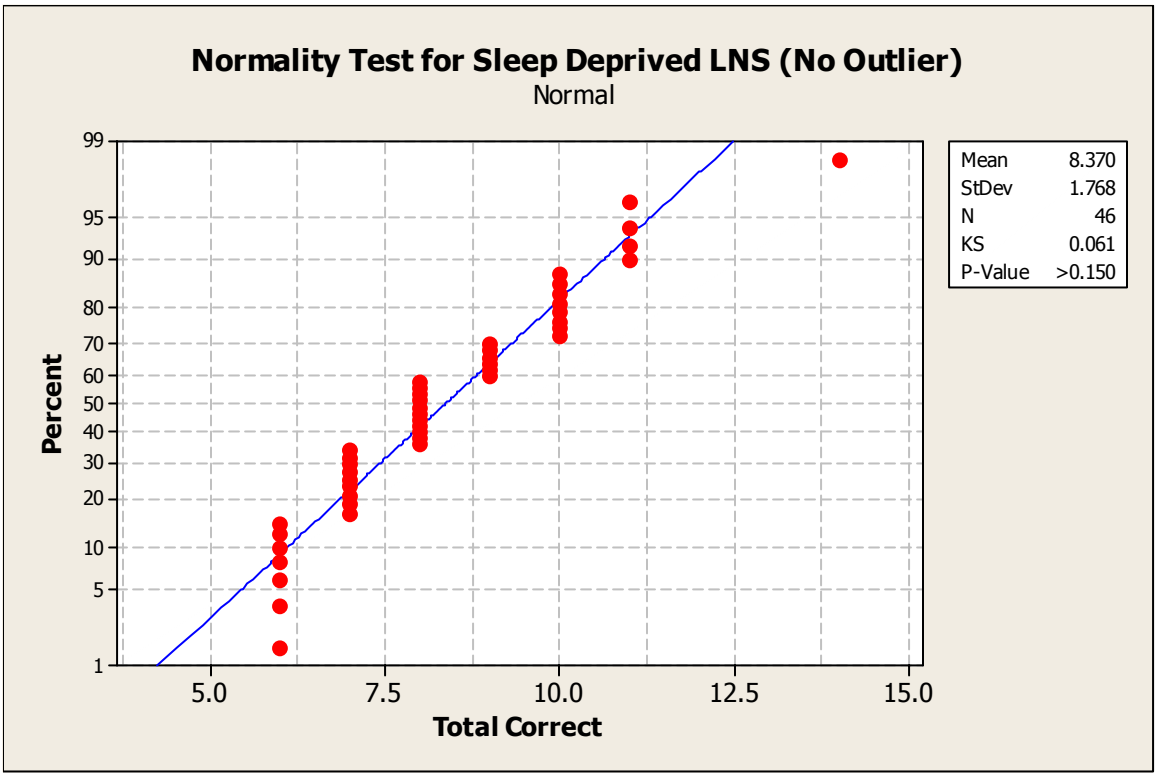


Figure D.3.5. Normality Test for Sleep-Deprived Letter-Number Sequencing with No Outliers

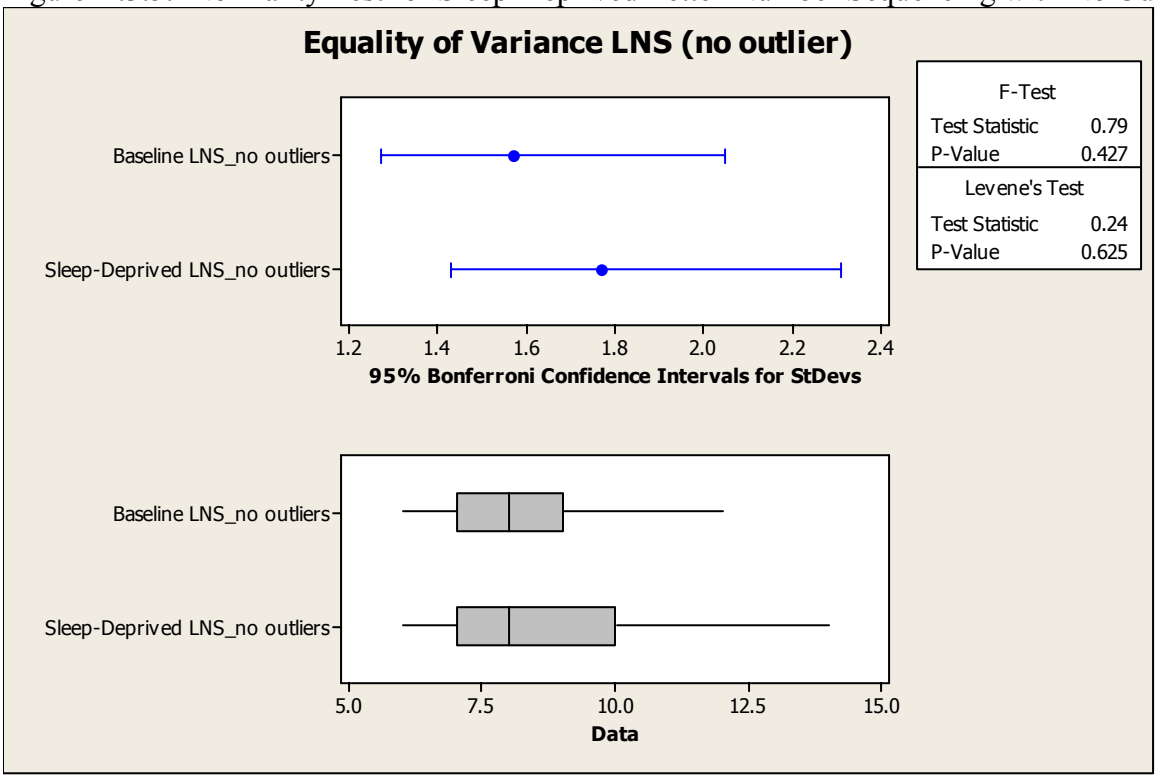


Figure D.3.6. Test for Equality of Variance for Letter Number Sequencing with No Outliers

D.4 LNS Improvement vs. Degradation

Table D.4.1 T-tests for Baselines Scores (Improved vs. Degraded vs. No Difference)

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Improved*	*No Diff*
Mean	7.611111111	7.111111111
Variance	1.545751634	0.611111111
Observations	18	9
Pooled Variance	1.246666667	
Hypothesized Mean Difference	0	
df	25	
t Stat	1.096908636	
P(T<=t) one-tail	0.141567564	
t Critical one-tail	1.708140745	
P(T<=t) two-tail	0.283135127	
t Critical two-tail	2.059538536	

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Degraded*	*Improved*
Mean	9.857142857	7.611111111
Variance	3.728571429	1.545751634
Observations	21	18
Pooled Variance	2.725654226	
Hypothesized Mean Difference	0	
df	37	
t Stat	4.235396822	
P(T<=t) one-tail	7.24931E-05	
t Critical one-tail	1.687093597	
P(T<=t) two-tail	0.000144986	
t Critical two-tail	2.026192447	

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Degraded*	*No Diff*
Mean	9.857142857	7.111111111
Variance	3.728571429	0.611111111
Observations	21	9
Pooled Variance	2.837868481	
Hypothesized Mean Difference	0	
Df	28	
t Stat	4.091473062	
P(T<=t) one-tail	0.000164327	
t Critical one-tail	1.701130908	
P(T<=t) two-tail	0.000328654	
t Critical two-tail	2.048407115	

D.4.1 Demographics and LNS Improvement vs. Degradation

Table D.4.1.1 ANOVA for Degraded vs. Improved and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
age	Between Groups	12.095	2	6.048	.469	.629
	Within Groups	580.571	45	12.902		
	Total	592.667	47			
IQ	Between Groups	228.190	2	114.095	.769	.470
	Within Groups	6678.476	45	148.411		
	Total	6906.667	47			
education	Between Groups	1.444	2	.722	.216	.806
	Within Groups	150.222	45	3.338		
	Total	151.667	47			
ME	Between Groups	95.417	2	47.708	.512	.603
	Within Groups	4194.500	45	93.211		
	Total	4289.917	47			

Table D.4.1.2 Chi-Square Tests for Degraded vs. Improved by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	8	1	0	9
	6.5625	0.9375	1.5	
	1.4375	0.0625	-1.5	
	16.67	2.08	0.00	18.75
	88.89	11.11	0.00	
	22.86	20.00	0.00	
1	12	1	5	18
	13.125	1.875	3	

	-1.125	-0.875	2	
	25.00	2.08	10.42	37.50
	66.67	5.56	27.78	
	34.29	20.00	62.50	
2	15	3	3	21
	15.313	2.1875	3.5	
	-0.313	0.8125	-0.5	
	31.25	6.25	6.25	43.75
	71.43	14.29	14.29	
	42.86	60.00	37.50	
Total	35	5	8	48
	72.92	10.42	16.67	100.00

Table D.4.1.3 Chi-Square Tests for Degraded vs. Improved by Language Cross Tabulation Matrix

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	4	4.0367	0.4011
Likelihood Ratio Chi-Square	4	5.3502	0.2532
Mantel-Haenszel Chi-Square	1	0.4733	0.4915
Phi Coefficient		0.2900	
Contingency Coefficient		0.2785	
Cramer's V		0.2051	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.4733	0.4915
2	Row Mean Scores Differ	2	2.5399	0.2809
3	General Association	4	3.9526	0.4125

Total Sample Size = 48

Table D.4.1.4 Chi-Square Tests for Degraded vs. Improved by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
0	3	6	9
	4.125	4.875	
	-1.125	1.125	
	6.25	12.50	18.75
	33.33	66.67	
	13.64	23.08	
1	7	11	18
	8.25	9.75	
	-1.25	1.25	
	14.58	22.92	37.50
	38.89	61.11	
	31.82	42.31	
2	12	9	21
	9.625	11.375	
	2.375	-2.375	
	25.00	18.75	43.75
	57.14	42.86	
	54.55	34.62	
Total	22	26	48
	45.83	54.17	100.00

Table D.4.1.5 Chi-Square Tests for Degraded vs. Improved by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	2	1.9980	0.3682
Likelihood Ratio Chi-Square	2	2.0121	0.3657
Mantel-Haenszel Chi-Square	1	1.7894	0.1810
Phi Coefficient		0.2040	
Contingency Coefficient		0.1999	
Cramer's V		0.2040	

WARNING: 33% of the cells have expected counts less
than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	1.7894	0.1810
2	Row Mean Scores Differ	2	1.9564	0.3760
3	General Association	2	1.9564	0.3760

Total Sample Size = 48

Table D.4.1.6 Chi-Square Tests for Degraded vs. Improved by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	2	1	1	5	9
	4.6875	1.3125	0.375	2.625	
	-2.688	-0.313	0.625	2.375	
	4.17	2.08	2.08	10.42	18.75
	22.22	11.11	11.11	55.56	
	8.00	14.29	50.00	35.71	
1	12	2	0	4	18
	9.375	2.625	0.75	5.25	
	2.625	-0.625	-0.75	-1.25	
	25.00	4.17	0.00	8.33	37.50
	66.67	11.11	0.00	22.22	
	48.00	28.57	0.00	28.57	
2	11	4	1	5	21
	10.938	3.0625	0.875	6.125	
	0.0625	0.9375	0.125	-1.125	
	22.92	8.33	2.08	10.42	43.75
	52.38	19.05	4.76	23.81	
	44.00	57.14	50.00	35.71	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.4.1.7 Chi-Square Tests for Degraded vs. Improved by Occupation

Statistic	DF	Value	Prob
Chi-Square	6	7.2490	0.2984
Likelihood Ratio Chi-Square	6	7.6153	0.2677
Mantel-Haenszel Chi-Square	1	2.2174	0.1365
Phi Coefficient		0.3886	
Contingency Coefficient		0.3622	
Cramer's V		0.2748	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.2174	0.1365
2	Row Mean Scores Differ	2	5.3383	0.0693
3	General Association	6	7.0980	0.3119

Total Sample Size = 48

Table D.4.1.8 Chi-Square Tests for Degraded vs. Improved by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	5	6	Total
0	1	6	2	0	0	9
	2.8125	4.3125	0.5625	0.1875	1.125	
	-1.813	1.6875	1.4375	-0.188	-1.125	
	2.08	12.50	4.17	0.00	0.00	18.75
	11.11	66.67	22.22	0.00	0.00	
	6.67	26.09	66.67	0.00	0.00	
1	5	9	1	0	3	18
	5.625	8.625	1.125	0.375	2.25	
	-0.625	0.375	-0.125	-0.375	0.75	
	10.42	18.75	2.08	0.00	6.25	37.50
	27.78	50.00	5.56	0.00	16.67	
	33.33	39.13	33.33	0.00	50.00	
2	9	8	0	1	3	21
	6.5625	10.063	1.3125	0.4375	2.625	
	2.4375	-2.063	-1.313	0.5625	0.375	
	18.75	16.67	0.00	2.08	6.25	43.75
	42.86	38.10	0.00	4.76	14.29	
	60.00	34.78	0.00	100.00	50.00	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.4.1.9 Chi-Square Tests for Degraded vs. Improved by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	8	10.9565	0.2042
Likelihood Ratio Chi-Square	8	12.5175	0.1296
Mantel-Haenszel Chi-Square	1	0.0313	0.8596
Phi Coefficient		0.4778	
Contingency Coefficient		0.4311	
Cramer's V		0.3378	

WARNING: 73% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0313	0.8596
2	Row Mean Scores Differ	2	0.1742	0.9166
3	General Association	8	10.7283	0.2176

Total Sample Size = 48

Table D.4.1.10 Chi-Square Tests for Degraded vs. Improved by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	3	2	4	0	9
	3.75	2.8125	2.25	0.1875	
	-0.75	-0.813	1.75	-0.188	
	6.25	4.17	8.33	0.00	18.75
	33.33	22.22	44.44	0.00	
	15.00	13.33	33.33	0.00	
1	9	6	3	0	18
	7.5	5.625	4.5	0.375	
	1.5	0.375	-1.5	-0.375	
	18.75	12.50	6.25	0.00	37.50
	50.00	33.33	16.67	0.00	
	45.00	40.00	25.00	0.00	
2	8	7	5	1	21
	8.75	6.5625	5.25	0.4375	
	-0.75	0.4375	-0.25	0.5625	
	16.67	14.58	10.42	2.08	43.75
	38.10	33.33	23.81	4.76	
	40.00	46.67	41.67	100.00	
Total	20	15	12	1	48
	41.67	31.25	25.00	2.08	100.00

Table D.4.1.11 Chi-Square Tests for Degraded vs. Improved by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	6	3.9619	0.6818
Likelihood Ratio Chi-Square	6	4.1590	0.6552
Mantel-Haenszel Chi-Square	1	0.0123	0.9115
Phi Coefficient		0.2873	
Contingency Coefficient		0.2761	
Cramer's V		0.2031	

WARNING: 58% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0123	0.9115
2	Row Mean Scores Differ	2	1.8783	0.3910
3	General Association	6	3.8794	0.6930

Total Sample Size = 48

D.5 Letter Sets Data (Normality and Equality of Variance)

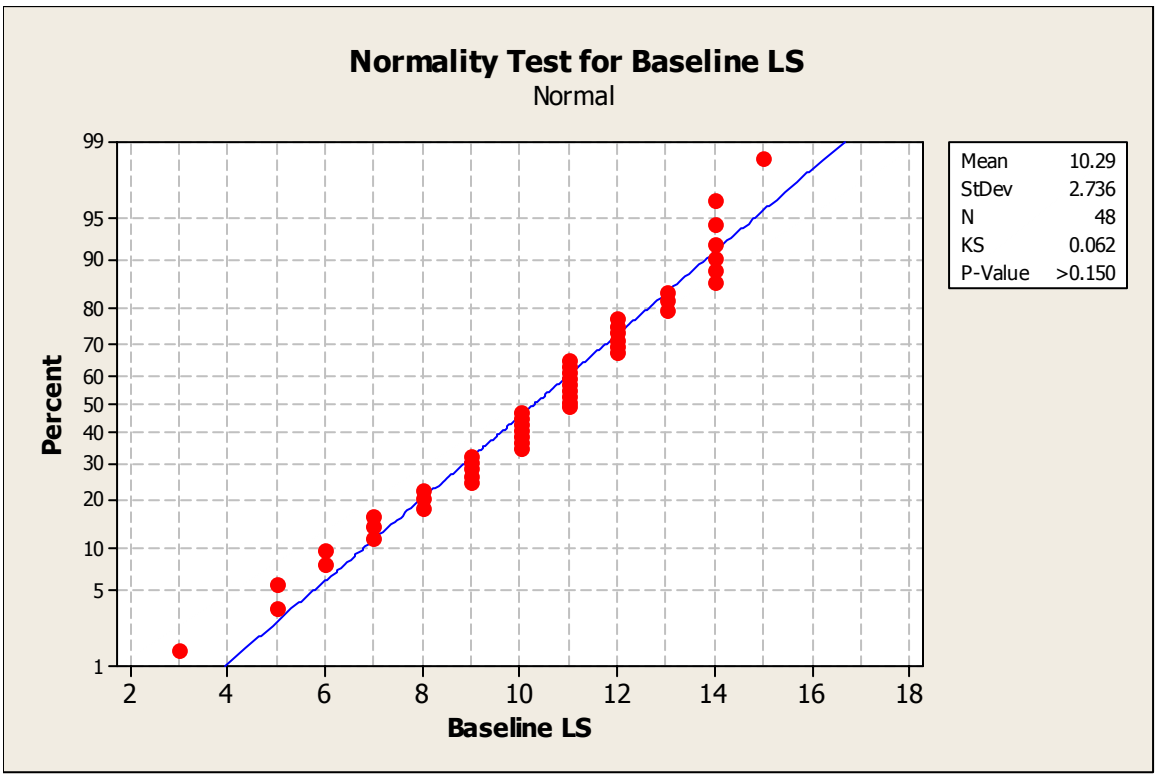


Figure D.5.1. Normality Test for Baseline Letter Sets Test

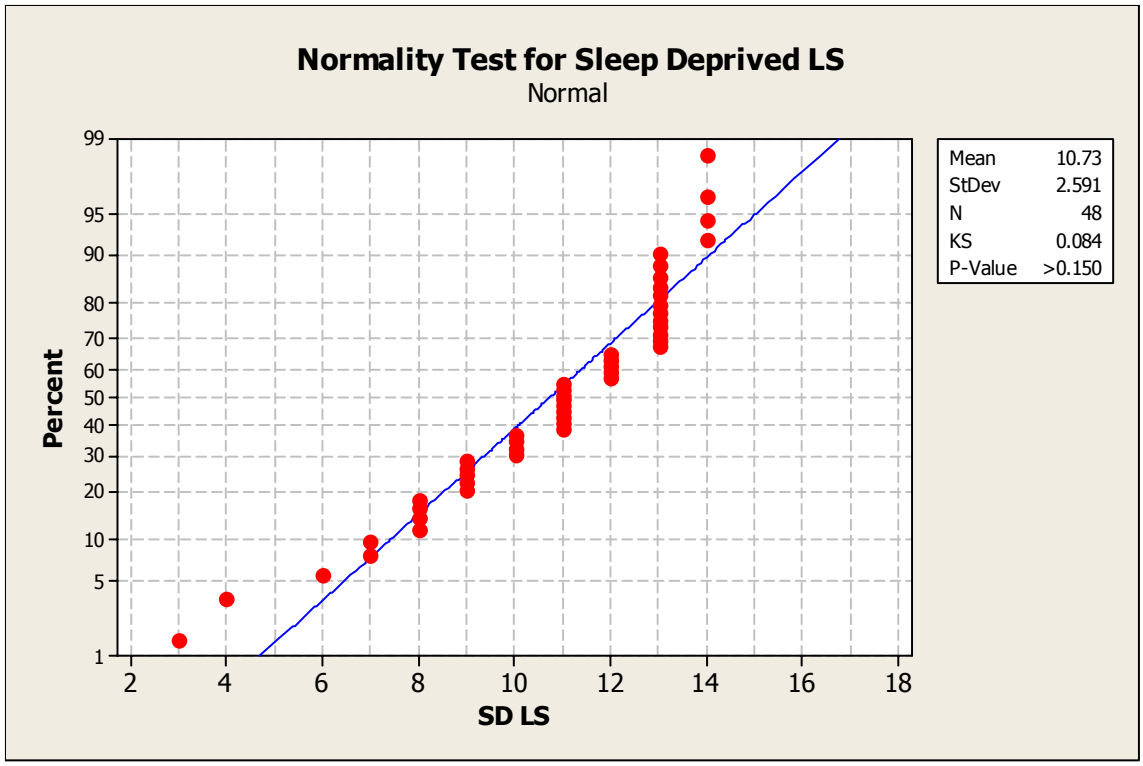


Figure D.5.2. Normality Test for Sleep-Deprived Letter Sets

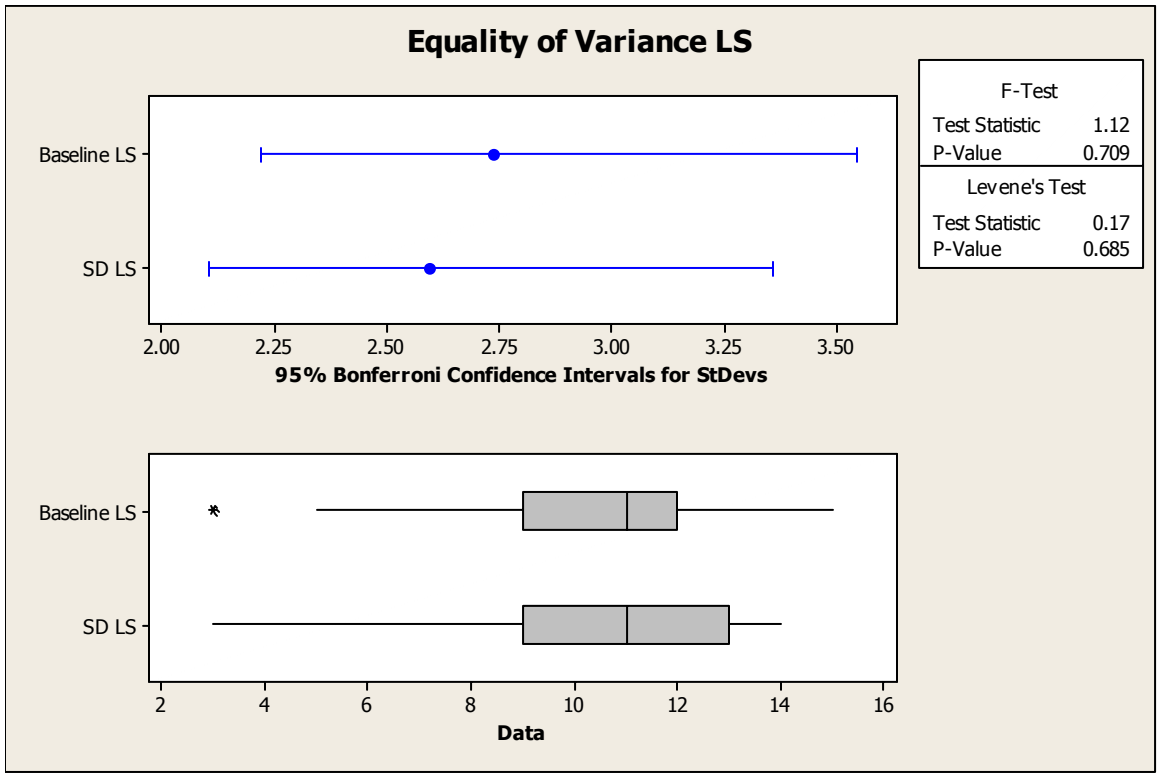


Figure D.5.3. Test for Equality of Variance for Letter Sets

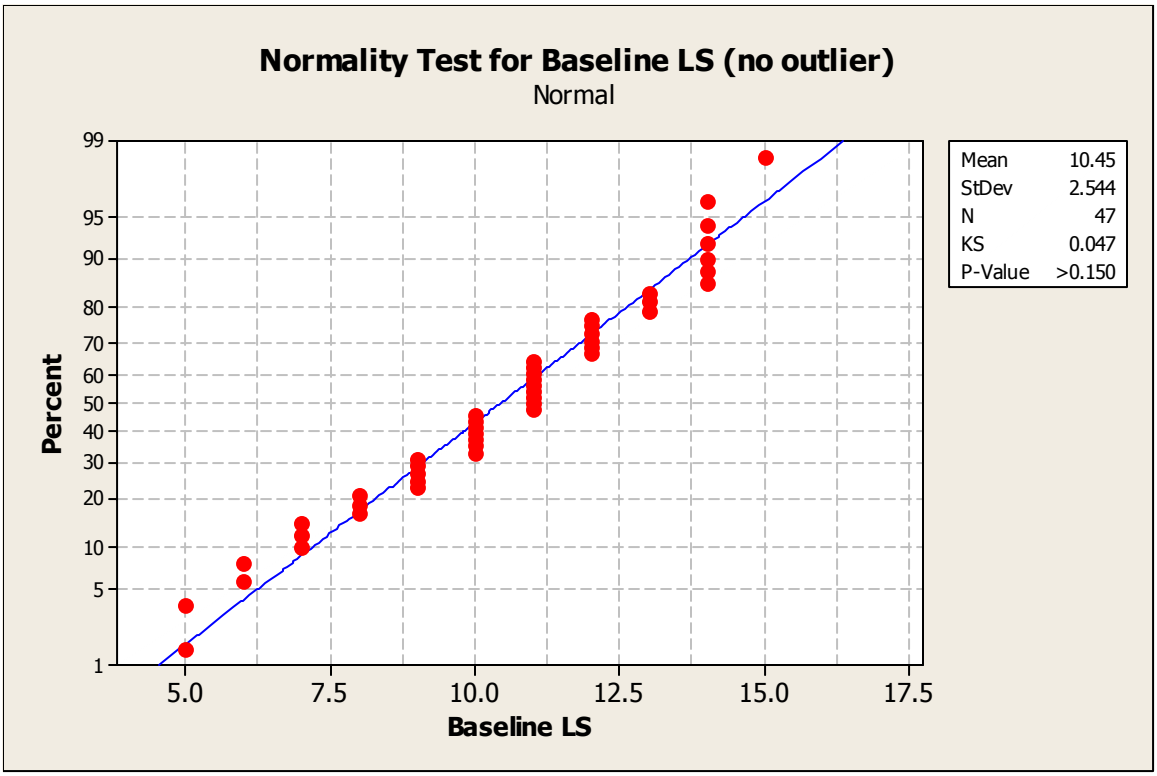


Figure D.5.4. Normality Test for Baseline Letter Sets with No Outliers

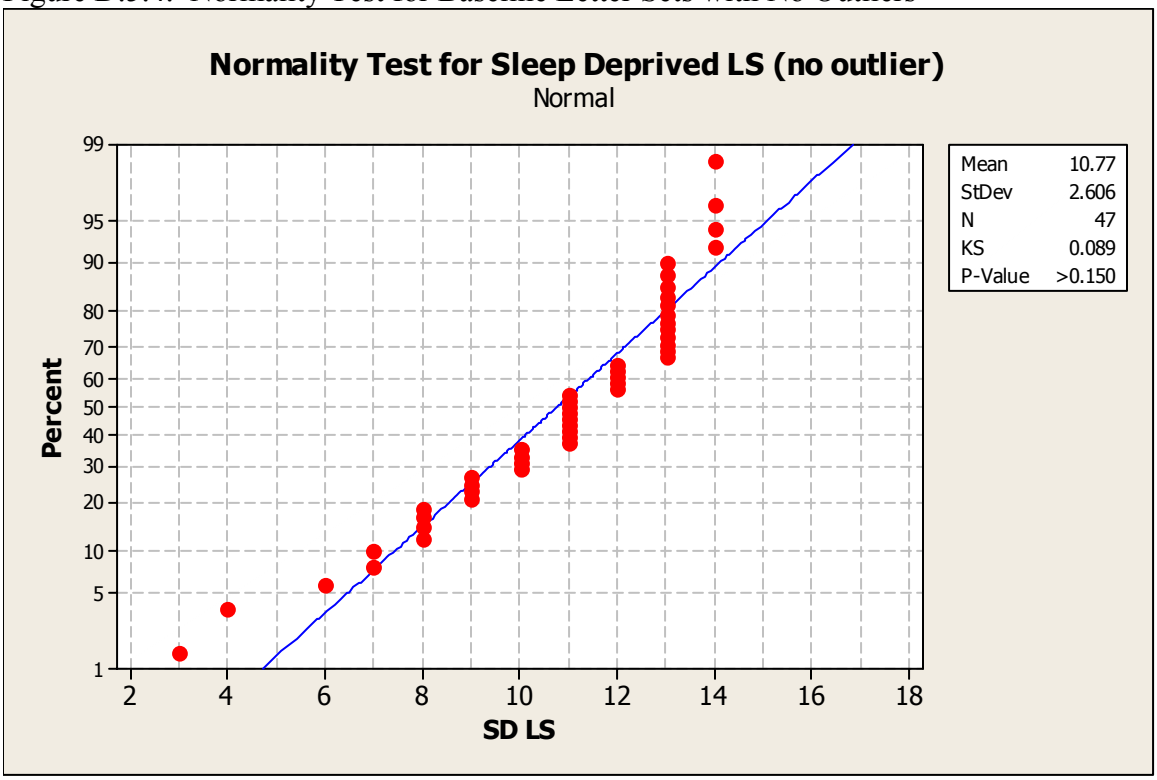


Figure D.5.5. Normality Test for Sleep-Deprived Letter Sets with No Outliers

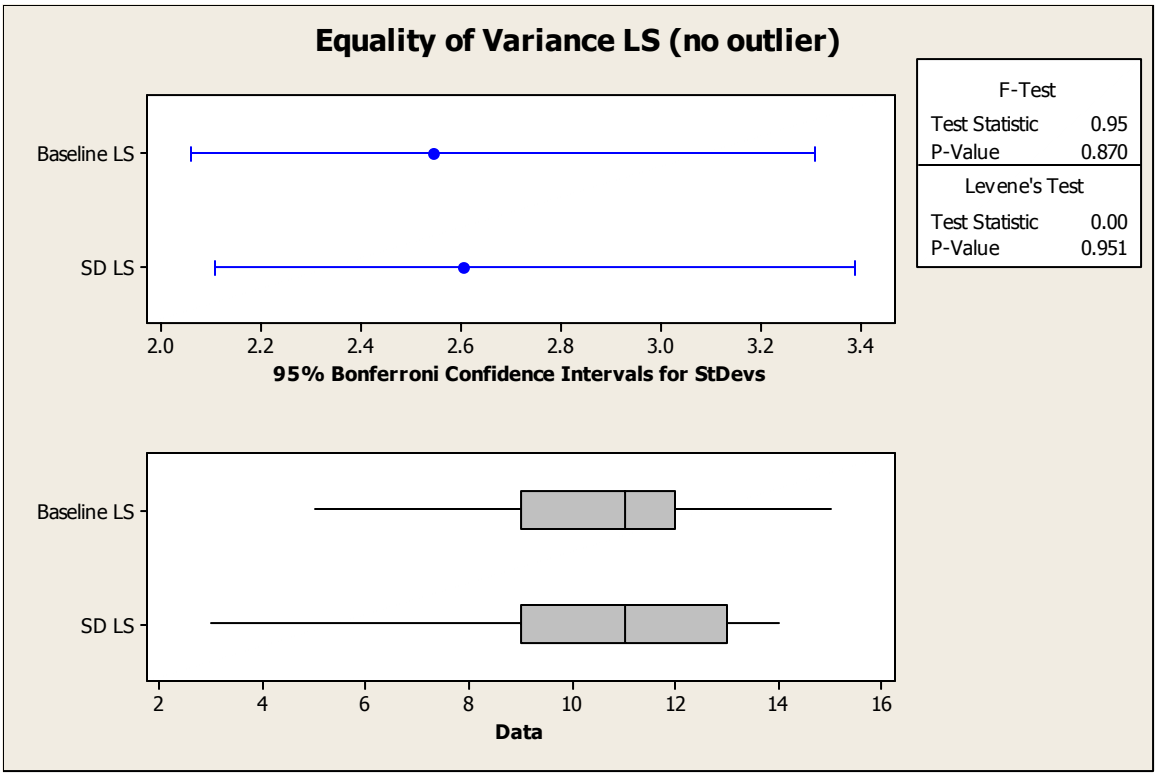


Figure D.5.6. Test for Equality of Variance for Letter Sets with No Outliers

D.6 LS Improvment vs. Degradation

Table D.6.1 T-test for Baseline Scores (Improved vs. Degraded vs. No difference)

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Improved*	*Degraded*
Mean	8.818181818	11.94736842
Variance	6.632034632	5.274853801
Observations	22	19
Pooled Variance	6.005643479	
Hypothesized Mean Difference	0	
Df	39	
	-	
t Stat	4.077067495	
P(T<=t) one-tail	0.000108671	
t Critical one-tail	1.684875122	
P(T<=t) two-tail	0.000217342	
t Critical two-tail	2.022690901	

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Improved*	*No Diff*

Mean	8.818181818	10.42857143
Variance	6.632034632	2.952380952
Observations	22	7
Pooled Variance	5.814333814	
Hypothesized Mean Difference	0	
Df	27	
t Stat	-1.539013185	
P(T<=t) one-tail	0.06771986	
t Critical one-tail	1.703288423	
P(T<=t) two-tail	0.135439721	
t Critical two-tail	2.051830493	

t-Test: Two-Sample Assuming Equal Variances

MEAN BASELINE SCORES	*Degraded*	*No diff*
Mean	11.94736842	10.42857143
Variance	5.274853801	2.952380952
Observations	19	7
Pooled Variance	4.694235589	
Hypothesized Mean Difference	0	
Df	24	
t Stat	1.585464178	
P(T<=t) one-tail	0.062975152	
t Critical one-tail	1.710882067	
P(T<=t) two-tail	0.125950305	
t Critical two-tail	2.063898547	

D.6.1 Demographics and Improvement Vs. Degradation

Table D.6.1.1 ANOVA for Improved vs. Degraded and Demogrpahics

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	1.108	2	.554	.042	.959
	Within Groups	591.559	45	13.146		
	Total	592.667	47			
IQ	Between Groups	185.125	2	92.562	.620	.543
	Within Groups	6721.542	45	149.368		
	Total	6906.667	47			
Education	Between Groups	4.341	2	2.170	.663	.520
	Within Groups	147.326	45	3.274		
	Total	151.667	47			
ME	Between Groups	366.679	2	183.340	2.103	.134
	Within Groups	3923.238	45	87.183		
	Total	4289.917	47			

Table D.6.1.2 Chi-Square tests for Degraded vs. Improved by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	7	0	0	7
	5.1042	0.7292	1.1667	
	1.8958	-0.729	-1.167	
	14.58	0.00	0.00	14.58
	100.00	0.00	0.00	
	20.00	0.00	0.00	
1	17	1	4	22
	16.042	2.2917	3.6667	
	0.9583	-1.292	0.3333	
	35.42	2.08	8.33	45.83
	77.27	4.55	18.18	
	48.57	20.00	50.00	
2	11	4	4	19
	13.854	1.9792	3.1667	
	-2.854	2.0208	0.8333	
	22.92	8.33	8.33	39.58
	57.89	21.05	21.05	
	31.43	80.00	50.00	
Total	35	5	8	48
	72.92	10.42	16.67	100.00

Table D.6.1.3 Chi-Square tests for Degraded vs. Improved by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	4	6.2863	0.1788
Likelihood Ratio Chi-Square	4	7.8550	0.0970
Mantel-Haenszel Chi-Square	1	3.3476	0.0673
Phi Coefficient		0.3619	
Contingency Coefficient		0.3403	
Cramer's V		0.2559	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	3.3476	0.0673
2	Row Mean Scores Differ	2	3.5036	0.1735
3	General Association	4	6.1553	0.1878

Total Sample Size = 48

Table D.6.1.4 Chi-Square tests for Degraded vs. Improved by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
0	4 3.2083 0.7917 8.33 57.14 18.18	3 3.7917 -0.792 6.25 42.86 11.54	7 14.58
1	10 10.083 -0.083 20.83 45.45 45.45	12 11.917 0.0833 25.00 54.55 46.15	22 45.83
2	8 8.7083 -0.708 16.67 42.11 36.36	11 10.292 0.7083 22.92 57.89 42.31	19 39.58
Total	22 45.83	26 54.17	48 100.00

Table D.6.1.5 Chi-Square tests for Degraded vs. Improved by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	2	0.4683	0.7913
Likelihood Ratio Chi-Square	2	0.4674	0.7916
Mantel-Haenszel Chi-Square	1	0.3858	0.5345
Phi Coefficient		0.0988	
Contingency Coefficient		0.0983	
Cramer's V		0.0988	

WARNING: 33% of the cells have expected counts less
than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.3858	0.5345
2	Row Mean Scores Differ	2	0.4585	0.7951
3	General Association	2	0.4585	0.7951

Total Sample Size = 48

Table D.6.1.6 Chi-Square tests for Degraded vs. Improved by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	3	3	0	1	7
	3.6458	1.0208	0.2917	2.0417	
	-0.646	1.9792	-0.292	-1.042	
	6.25	6.25	0.00	2.08	14.58
	42.86	42.86	0.00	14.29	
	12.00	42.86	0.00	7.14	
1	12	2	0	8	22
	11.458	3.2083	0.9167	6.4167	
	0.5417	-1.208	-0.917	1.5833	
	25.00	4.17	0.00	16.67	45.83
	54.55	9.09	0.00	36.36	
	48.00	28.57	0.00	57.14	
2	10	2	2	5	19
	9.8958	2.7708	0.7917	5.5417	
	0.1042	-0.771	1.2083	-0.542	
	20.83	4.17	4.17	10.42	39.58
	52.63	10.53	10.53	26.32	
	40.00	28.57	100.00	35.71	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.6.1.7 Chi-Square tests for Degraded vs. Improved by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	6	8.6755	0.1927
Likelihood Ratio Chi-Square	6	8.2013	0.2237
Mantel-Haenszel Chi-Square	1	0.0759	0.7830
Phi Coefficient		0.4251	
Contingency Coefficient		0.3912	
Cramer's V		0.3006	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0759	0.7830
2	Row Mean Scores Differ	2	0.3190	0.8526
3	General Association	6	8.4948	0.2040

Total Sample Size = 48

Table D.6.1.8 Chi-Square tests for Degraded vs. Improved by Ethnicity Cross Tabulation Matrix

Table of imp1deg2 by ethnicity

imp1deg2(imp1deg2) ethnicity(ethnicity)

Frequency / Expected / Deviation / Percent / Row Pct / Col Pct	1	2	4	5	6	Total
0	1	5	1	0	0	7
	2.1875	3.3542	0.4375	0.1458	0.875	
	-1.188	1.6458	0.5625	-0.146	-0.875	
	2.08	10.42	2.08	0.00	0.00	14.58
	14.29	71.43	14.29	0.00	0.00	
	6.67	21.74	33.33	0.00	0.00	
1	7	11	0	0	4	22
	6.875	10.542	1.375	0.4583	2.75	
	0.125	0.4583	-1.375	-0.458	1.25	
	14.58	22.92	0.00	0.00	8.33	45.83
	31.82	50.00	0.00	0.00	18.18	
	46.67	47.83	0.00	0.00	66.67	
2	7	7	2	1	2	19
	5.9375	9.1042	1.1875	0.3958	2.375	
	1.0625	-2.104	0.8125	0.6042	-0.375	
	14.58	14.58	4.17	2.08	4.17	39.58
	36.84	36.84	10.53	5.26	10.53	
	46.67	30.43	66.67	100.00	33.33	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.6.1.9 Chi-Square tests for Degraded vs. Improved by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	8	7.8337	0.4499
Likelihood Ratio Chi-Square	8	10.1428	0.2551
Mantel-Haenszel Chi-Square	1	0.1020	0.7495
Phi Coefficient		0.4040	
Contingency Coefficient		0.3746	
Cramer's V		0.2857	

WARNING: 73% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1020	0.7495
2	Row Mean Scores Differ	2	0.1663	0.9202
3	General Association	8	7.6705	0.4663

Total Sample Size = 48

Table D.6.1.10 Chi-Square tests for Degraded vs. Improved by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	3 2.0417 0.9583 6.25 42.86 21.43	4 3.9375 0.0625 8.33 57.14 14.81	0 1.0208 -1.021 0.00 0.00 0.00	7 14.58
1	7 6.4167 0.5833 14.58 31.82 50.00	12 12.375 -0.375 25.00 54.55 44.44	3 3.2083 -0.208 6.25 13.64 42.86	22 45.83
2	4 5.5417 -1.542 8.33 21.05 28.57	11 10.688 0.3125 22.92 57.89 40.74	4 2.7708 1.2292 8.33 21.05 57.14	19 39.58
Total	14 29.17	27 56.25	7 14.58	48 100.00

Table D.6.1.11 Chi-Square tests for Degraded vs. Improved by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	4	2.5329	0.6388
Likelihood Ratio Chi-Square	4	3.4751	0.4817
Mantel-Haenszel Chi-Square	1	2.3077	0.1287
Phi Coefficient		0.2297	
Contingency Coefficient		0.2239	
Cramer's V		0.1624	

WARNING: 56% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.3077	0.1287
2	Row Mean Scores Differ	2	2.3340	0.3113
3	General Association	4	2.4801	0.6482

Total Sample Size = 48

D.7 Nonsense Syllogisms Data (Normality and Equality of Variance)

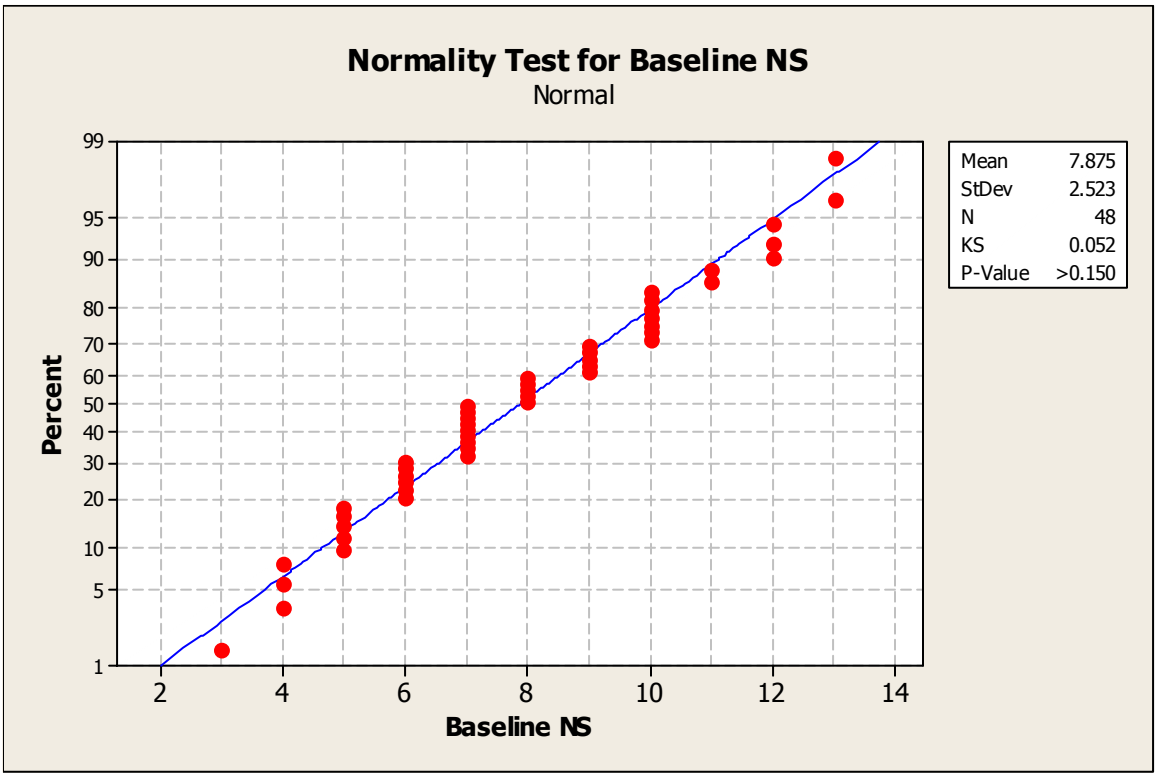


Figure D.7.1. Normality Test for Baseline Nonsense Syllogisms Test

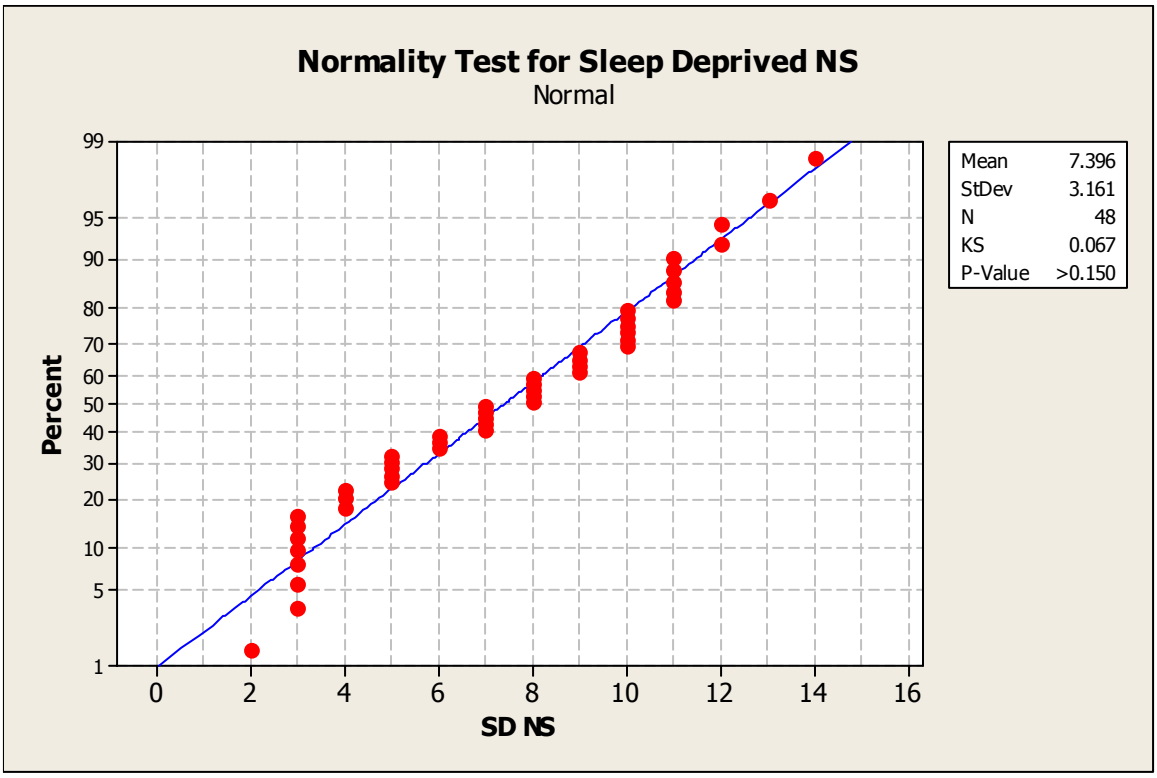


Figure D.7.2. Normality Test for Sleep Deprived Nonsense Syllogisms Test

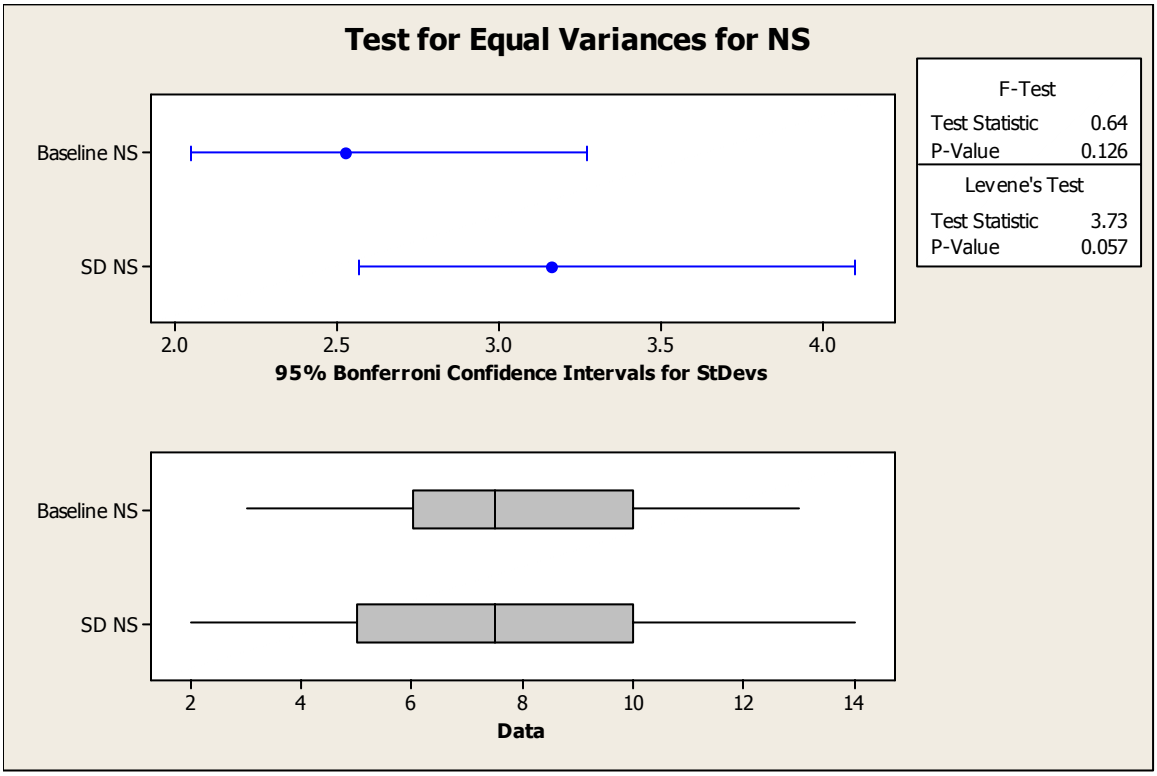


Figure D.7.3. Test for Equality of Variance for Nonsense Syllogisms

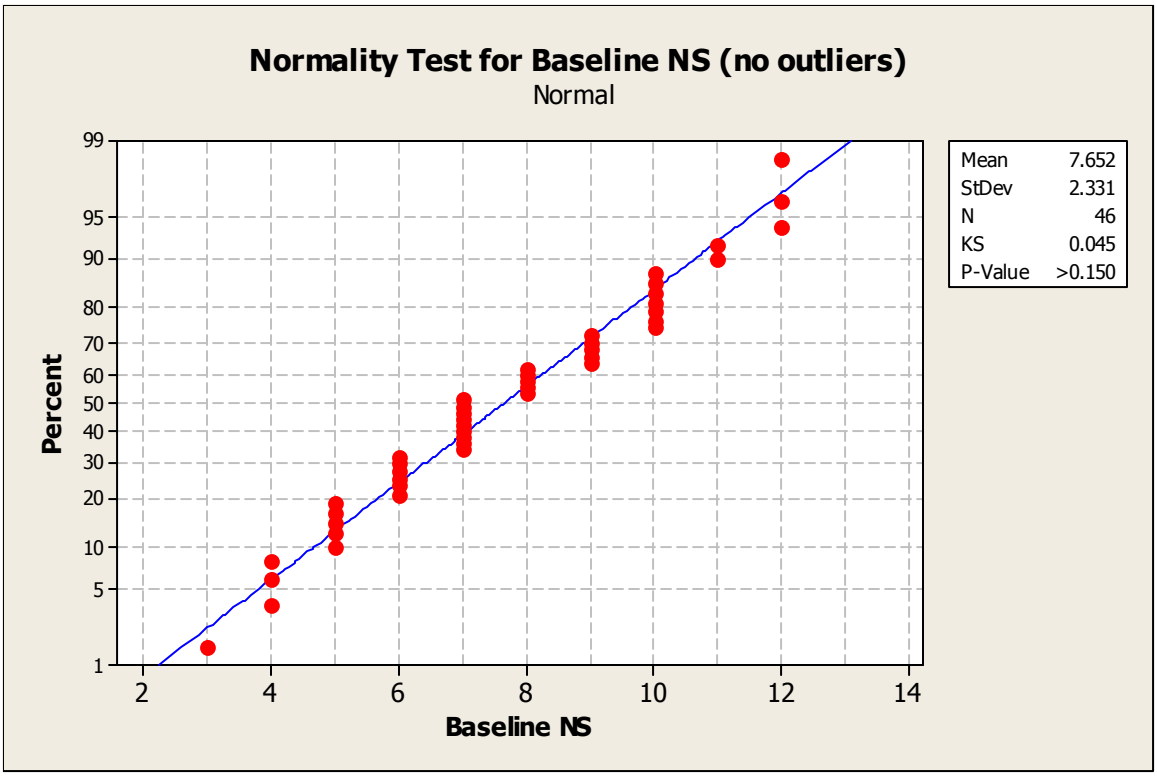


Figure D.7.4. Normality Test for Baseline Nonsense Syllogisms Test (no outliers)

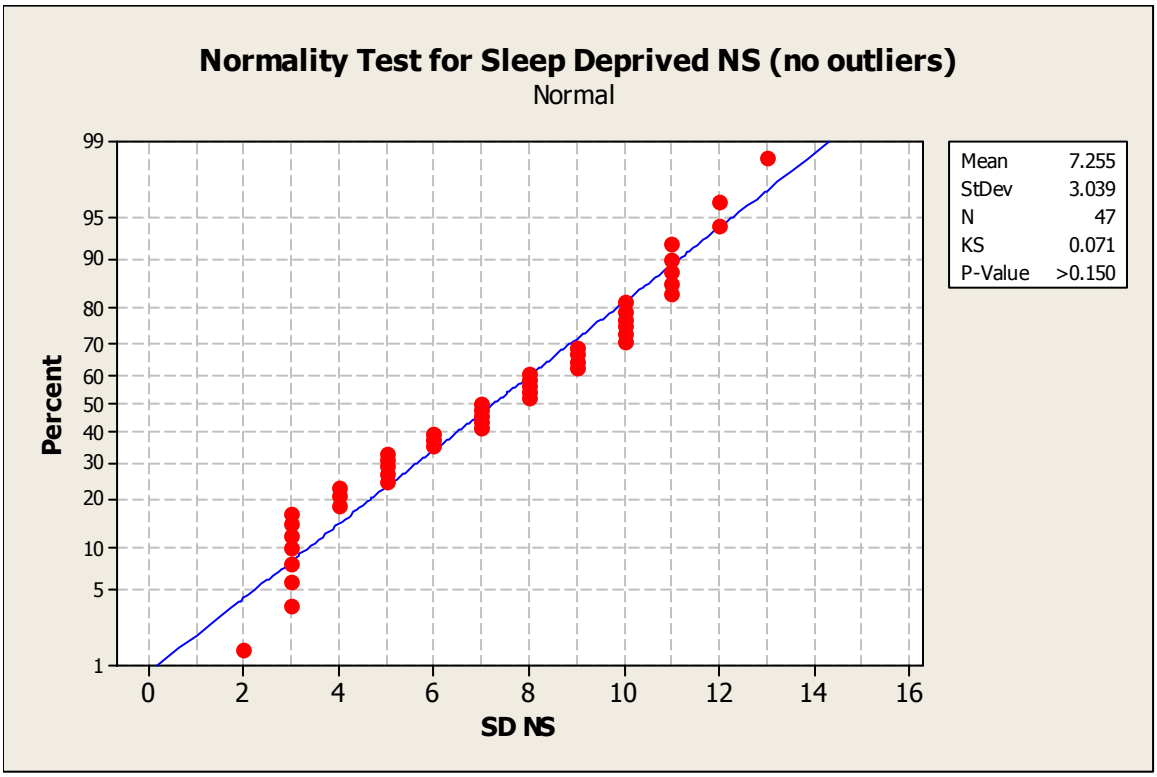
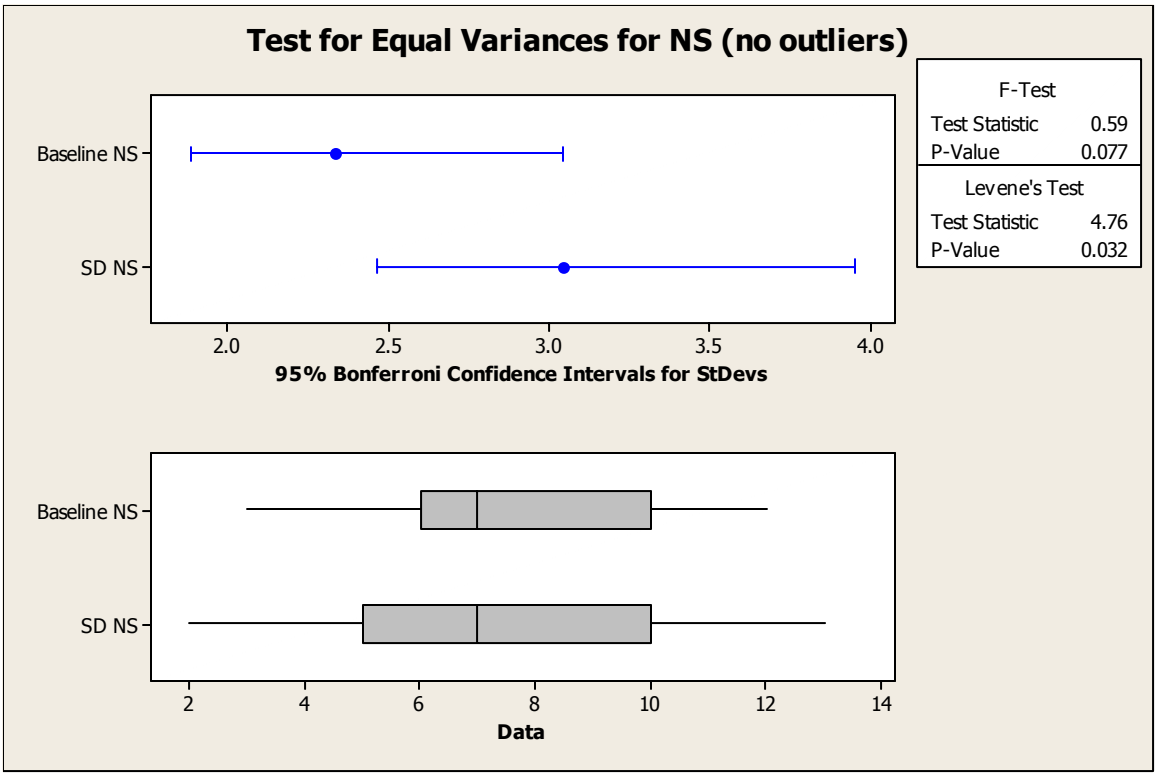


Figure D.7.5. Normality Test for Sleep Deprived Nonsense Syllogisms Test (no outliers)



D.7.6 Test for Equality of Variance for Nonsense Syllogisms (no outliers)

D.8 NS Improvement vs. Degradation

Table D.8.1 T-test for Baselines Score (Improved vs. Degraded vs. No Difference)

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Improved*	*Degraded*
Mean	7.473684211	8.333333333
Variance	6.707602339	6.405797101
Observations	19	24
Hypothesized Mean Difference	0	
Df	38	
	-	
t Stat	1.091808384	
P(T<=t) one-tail	0.140895753	
t Critical one-tail	1.685954461	
P(T<=t) two-tail	0.281791506	
t Critical two-tail	2.024394147	

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Improved*	*No Diff*
Mean	7.473684211	7.2
Variance	6.707602339	5.2

Observations	19	5
Hypothesized Mean Difference	0	
Df	7	
t Stat	0.231883177	
P(T<=t) one-tail	0.411630861	
t Critical one-tail	1.894578604	
P(T<=t) two-tail	0.823261723	
t Critical two-tail	2.364624251	

MEAN BASELINE SCORES
t-Test: Two-Sample Assuming Unequal Variances

	Degraded	*No Diff*
Mean	8.333333333	7.2
Variance	6.405797101	5.2
Observations	24	5
Hypothesized Mean Difference	0	
Df	6	
t Stat	0.991368508	
P(T<=t) one-tail	0.179893212	
t Critical one-tail	1.943180274	
P(T<=t) two-tail	0.359786424	
t Critical two-tail	2.446911846	

D.8.1 Demographics and Improvement vs. Degradation

Table D.8.1.1 ANOVA for Improved vs. Degraded and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	2.912	2	1.456	.111	.895
	Within Groups	589.754	45	13.106		
	Total	592.667	47			
IQ	Between Groups	888.919	2	444.460	3.324	.045
	Within Groups	6017.747	45	133.728		
	Total	6906.667	47			
Education	Between Groups	8.719	2	4.359	1.372	.264
	Within Groups	142.948	45	3.177		
	Total	151.667	47			
ME	Between Groups	361.232	2	180.616	2.069	.138
	Within Groups	3928.685	45	87.304		
	Total	4289.917	47			

Table D.8.1.2 Chi Square Tests for Improved vs. Degraded by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	5	0	0	5
	3.6458	0.5208	0.8333	
	1.3542	-0.521	-0.833	
	10.42	0.00	0.00	10.42
	100.00	0.00	0.00	
	14.29	0.00	0.00	
1	13	3	3	19
	13.854	1.9792	3.1667	
	-0.854	1.0208	-0.167	
	27.08	6.25	6.25	39.58
	68.42	15.79	15.79	
	37.14	60.00	37.50	
2	17	2	5	24
	17.5	2.5	4	
	-0.5	-0.5	1	
	35.42	4.17	10.42	50.00
	70.83	8.33	20.83	
	48.57	40.00	62.50	
Total	35	5	8	48
	72.92	10.42	16.67	100.00

Table D.8.1.3 Chi Square Tests for Improved vs. Degraded by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	4	2.8094	0.5902
Likelihood Ratio Chi-Square	4	4.0285	0.4022
Mantel-Haenszel Chi-Square	1	1.0698	0.3010
Phi Coefficient		0.2419	
Contingency Coefficient		0.2351	
Cramer's V		0.1711	

WARNING: 78% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob

1	Nonzero Correlation	1	1.0698	0.3010
2	Row Mean Scores Differ	2	1.8177	0.4030
3	General Association	4	2.7509	0.6003

Total Sample Size = 48

Table D.8.1.4 Chi Square Tests for Improved vs. Degraded by Gender Cross Tabulation Matrix)

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
0	3	2	5
	2.2917	2.7083	
	0.7083	-0.708	
	6.25	4.17	10.42
	60.00	40.00	
	13.64	7.69	
1	8	11	19
	8.7083	10.292	
	-0.708	0.7083	
	16.67	22.92	39.58
	42.11	57.89	
	36.36	42.31	
2	11	13	24
	11	13	
	0	0	
	22.92	27.08	50.00
	45.83	54.17	
	50.00	50.00	
Total	22	26	48
	45.83	54.17	100.00

Table D.8.1.5 Chi Square Tests for Improved vs. Degraded by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	2	0.5106	0.7747
Likelihood Ratio Chi-Square	2	0.5102	0.7749
Mantel-Haenszel Chi-Square	1	0.0921	0.7615
Phi Coefficient		0.1031	
Contingency Coefficient		0.1026	
Cramer's V		0.1031	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0921	0.7615
2	Row Mean Scores Differ	2	0.4999	0.7788
3	General Association	2	0.4999	0.7788

Total Sample Size = 48

Table D.8.1.6 Chi Square Tests for Improved vs. Degraded by Occupation Cross Tabulation Matrix

requency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	1	3	0	1	5
	2.6042	0.7292	0.2083	1.4583	
	-1.604	2.2708	-0.208	-0.458	
	2.08	6.25	0.00	2.08	10.42
	20.00	60.00	0.00	20.00	
	4.00	42.86	0.00	7.14	
1	14	2	0	3	19
	9.8958	2.7708	0.7917	5.5417	
	4.1042	-0.771	-0.792	-2.542	
	29.17	4.17	0.00	6.25	39.58
	73.68	10.53	0.00	15.79	
	56.00	28.57	0.00	21.43	
2	10	2	2	10	24
	12.5	3.5	1	7	
	-2.5	-1.5	1	3	
	20.83	4.17	4.17	20.83	50.00
	41.67	8.33	8.33	41.67	
	40.00	28.57	100.00	71.43	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.8.1.7 Chi Square Tests for Improved vs. Degraded by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	6	15.7151	0.0154
Likelihood Ratio Chi-Square	6	13.7510	0.0325
Mantel-Haenszel Chi-Square	1	2.1589	0.1417
Phi Coefficient		0.5722	
Contingency Coefficient		0.4966	

Cramer's V	0.4046

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.1589	0.1417
2	Row Mean Scores Differ	2	5.1557	0.0759
3	General Association	6	15.3877	0.0174

Total Sample Size = 48

Table D.8.1.8 Chi Square Tests for Improved vs. Degraded by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	5	6	Total
0	2	2	0	0	1	5
	1.5625	2.3958	0.3125	0.1042	0.625	
	0.4375	-0.396	-0.313	-0.104	0.375	
	4.17	4.17	0.00	0.00	2.08	10.42
	40.00	40.00	0.00	0.00	20.00	
	13.33	8.70	0.00	0.00	16.67	
1	10	7	0	0	2	19
	5.9375	9.1042	1.1875	0.3958	2.375	
	4.0625	-2.104	-1.188	-0.396	-0.375	
	20.83	14.58	0.00	0.00	4.17	39.58
	52.63	36.84	0.00	0.00	10.53	
	66.67	30.43	0.00	0.00	33.33	
2	3	14	3	1	3	24
	7.5	11.5	1.5	0.5	3	
	-4.5	2.5	1.5	0.5	0	
	6.25	29.17	6.25	2.08	6.25	50.00
	12.50	58.33	12.50	4.17	12.50	
	20.00	60.87	100.00	100.00	50.00	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.8.1.9 Chi Square Tests for Improved vs. Degraded by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	8	10.9815	0.2027
Likelihood Ratio Chi-Square	8	12.8195	0.1182
Mantel-Haenszel Chi-Square	1	1.3761	0.2408
Phi Coefficient		0.4783	
Contingency Coefficient		0.4315	
Cramer's V		0.3382	

WARNING: 73% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	1.3761	0.2408
2	Row Mean Scores Differ	2	2.9121	0.2332
3	General Association	8	10.7527	0.2161

Total Sample Size = 48

D.8.1.10 Chi Square Tests for Improved vs. Degraded by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	4	0	1	5
	1.4583	2.8125	0.7292	
	2.5417	-2.813	0.2708	
	8.33	0.00	2.08	10.42
	80.00	0.00	20.00	
	28.57	0.00	14.29	
1	7	8	4	19
	5.5417	10.688	2.7708	
	1.4583	-2.688	1.2292	
	14.58	16.67	8.33	39.58
	36.84	42.11	21.05	
	50.00	29.63	57.14	
2	3	19	2	24
	7	13.5	3.5	
	-4	5.5	-1.5	
	6.25	39.58	4.17	50.00
	12.50	79.17	8.33	
	21.43	70.37	28.57	
Total	14	27	7	48
	29.17	56.25	14.58	100.00

D.8.1.11 Chi Square Tests for Improved vs. Degraded by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	4	14.1170	0.0069
Likelihood Ratio Chi-Square	4	15.9416	0.0031
Mantel-Haenszel Chi-Square	1	2.4928	0.1144
Phi Coefficient		0.5423	
Contingency Coefficient		0.4767	
Cramer's V		0.3835	

WARNING: 56% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.4928	0.1144
2	Row Mean Scores Differ	2	3.0453	0.2181
3	General Association	4	13.8229	0.0079

Total Sample Size = 48

D.8.1.1 Post Hoc WASI Analysis

Table D.8.1.1.1 T-tests for Improved vs. Degraded based on WASI Score Group Statistics

	Imp/Deg	N	Mean	Std. Deviation	Std. Error Mean
IQ	Improve	19	111.9474	13.71728	3.14696
	Degrade	24	103.5000	9.84665	2.00994

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IQ	Equal variances assumed	3.665	.063	2.350	41	.024	8.44737	3.59426	1.18860	15.7061
	Equal variances not assumed			2.262	31.569	.031	8.44737	3.73406	.83726	16.0574

Table D.8.1.1.2 T-tests for Improved vs. No Difference based on WASI Score Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IQ	Equal variances assumed	1.586	.221	1.539	22	.138	10.14737	6.59512	-3.5300	23.8248
	Equal variances not assumed			1.854	8.471	.099	10.14737	5.47205	-2.3502	22.6449

Table D.8.1.1.3 T-tests for No Difference vs. Degraded based on WASI Score Group Statistics

	Imp/Deg	N	Mean	Std. Deviation	Std. Error Mean
IQ	No Diff	5	101.8000	10.01000	4.47661
	Deg	24	103.5000	9.84665	2.00994

Table D.8.1.1.4 T-tests for No Difference vs. Degraded based on WASI Score Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
IQ	Equal variances assumed	.033	.857	-.350	27	.729	-1.70000	4.85255	-11.6566	8.25662
	Equal variances not assumed			-.346	5.735	.741	-1.70000	4.90712	-13.8432	10.4432

Table D.8.1.1.5 WASI ANOVA with Language and Education Covariates Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1210.110(a)	4	302.528	2.284	.076
Intercept	4123.993	1	4123.993	31.130	.000
lang	109.381	1	109.381	.826	.369
education	316.222	1	316.222	2.387	.130
imp1deg2	1141.208	2	570.604	4.307	.020
Error	5696.556	43	132.478		
Total	553040.000	48			
Corrected Total	6906.667	47			

a R Squared = .175 (Adjusted R Squared = .098)

Table D.8.1.1.6 WASI ANOVA with Education Covariate Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1100.730(a)	3	366.910	2.781	.052
Intercept	4945.452	1	4945.452	37.479	.000
education	211.810	1	211.810	1.605	.212
imp1deg2	1046.927	2	523.463	3.967	.026
Error	5805.937	44	131.953		
Total	553040.000	48			
Corrected Total	6906.667	47			

a R Squared = .159 (Adjusted R Squared = .102)

Table D.8.1.1.7 WASI ANOVA with Language Covariate Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	893.888(a)	3	297.963	2.180	.104
Intercept	114971.855	1	114971.855	841.335	.000
Lang	4.969	1	4.969	.036	.850
imp1deg2	893.744	2	446.872	3.270	.047
Error	6012.778	44	136.654		
Total	553040.000	48			
Corrected Total	6906.667	47			

a R Squared = .129 (Adjusted R Squared = .070)

D.9 Maze Tracing Test Data (Normality and Equality of Variance)

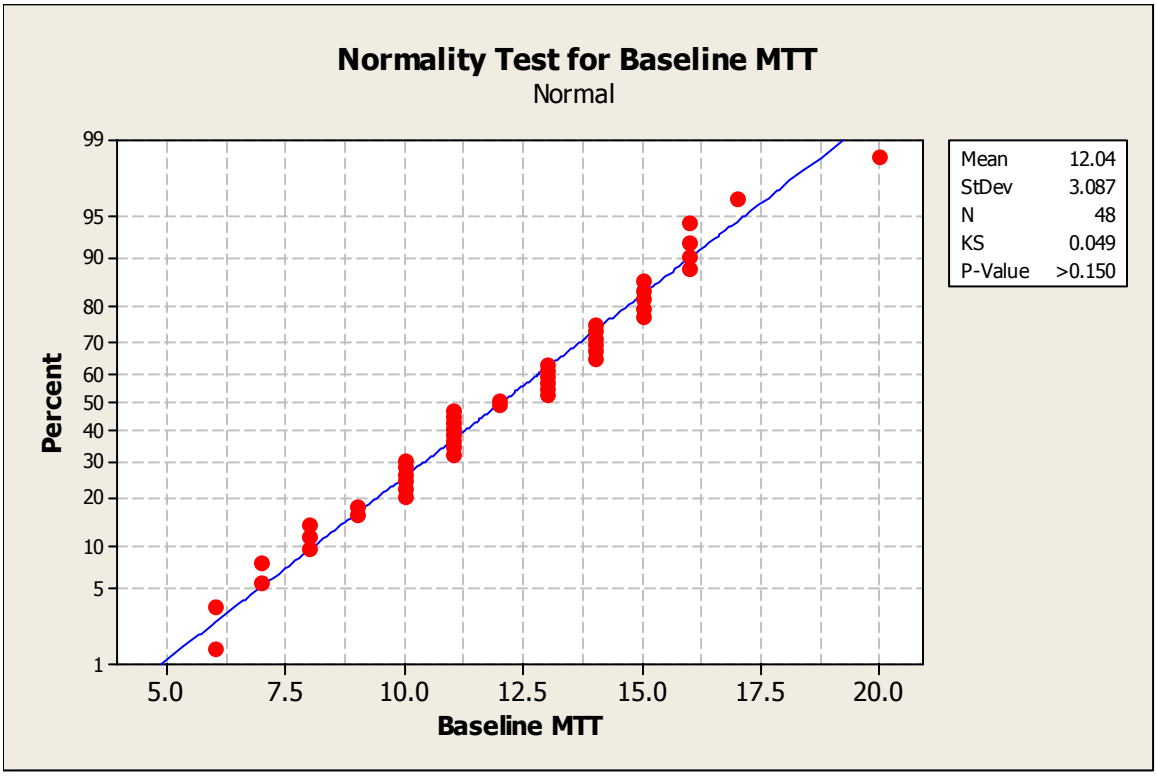


Figure D.9.1 Normality Test for Baseline Maze Tracing Test

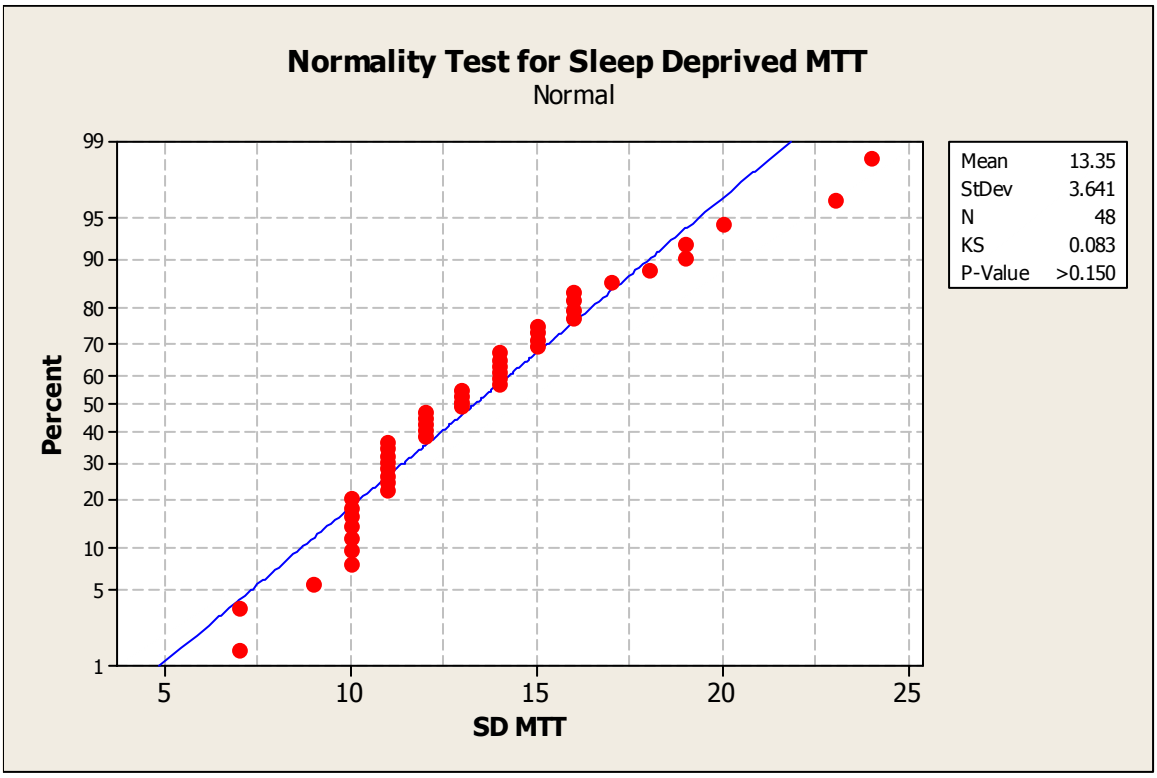


Figure D.9.2 Normality Test for Sleep-Deprived Maze Tracing Test

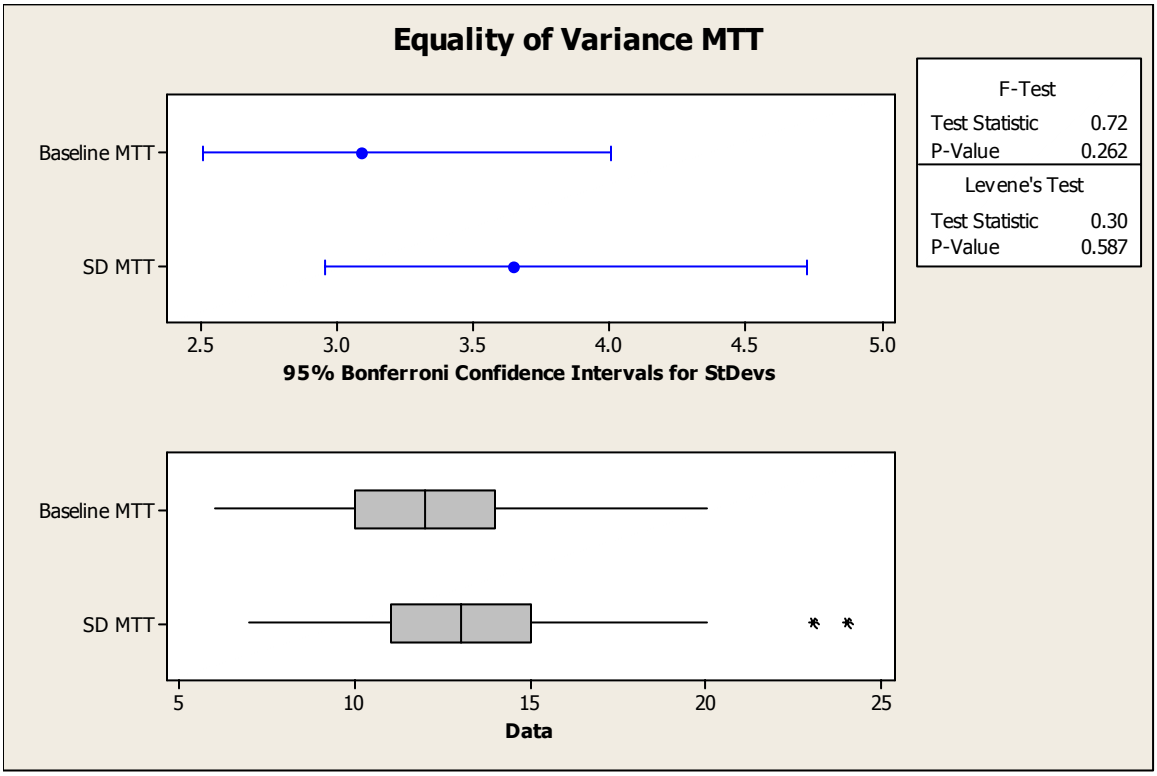


Figure D.9.3 Equality of Variance for Maze Tracing Test

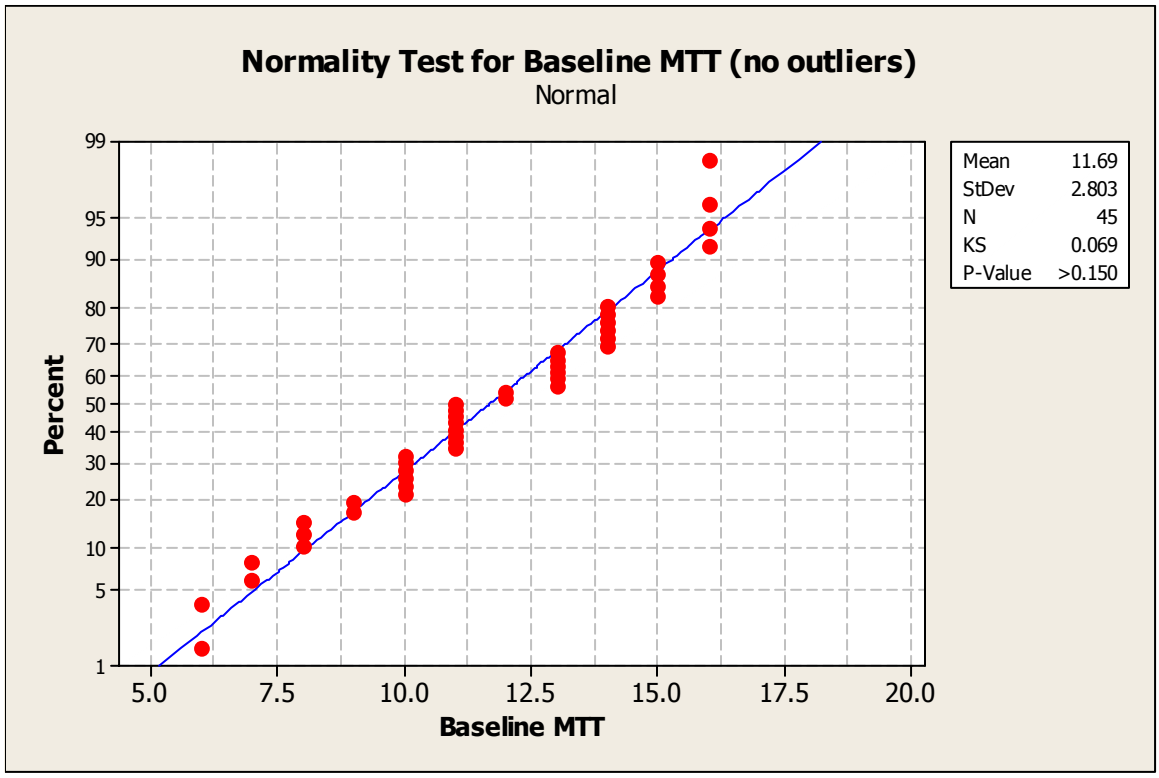


Figure D.9.4 Normality Test for Baseline Maze Tracing Test with No Outliers

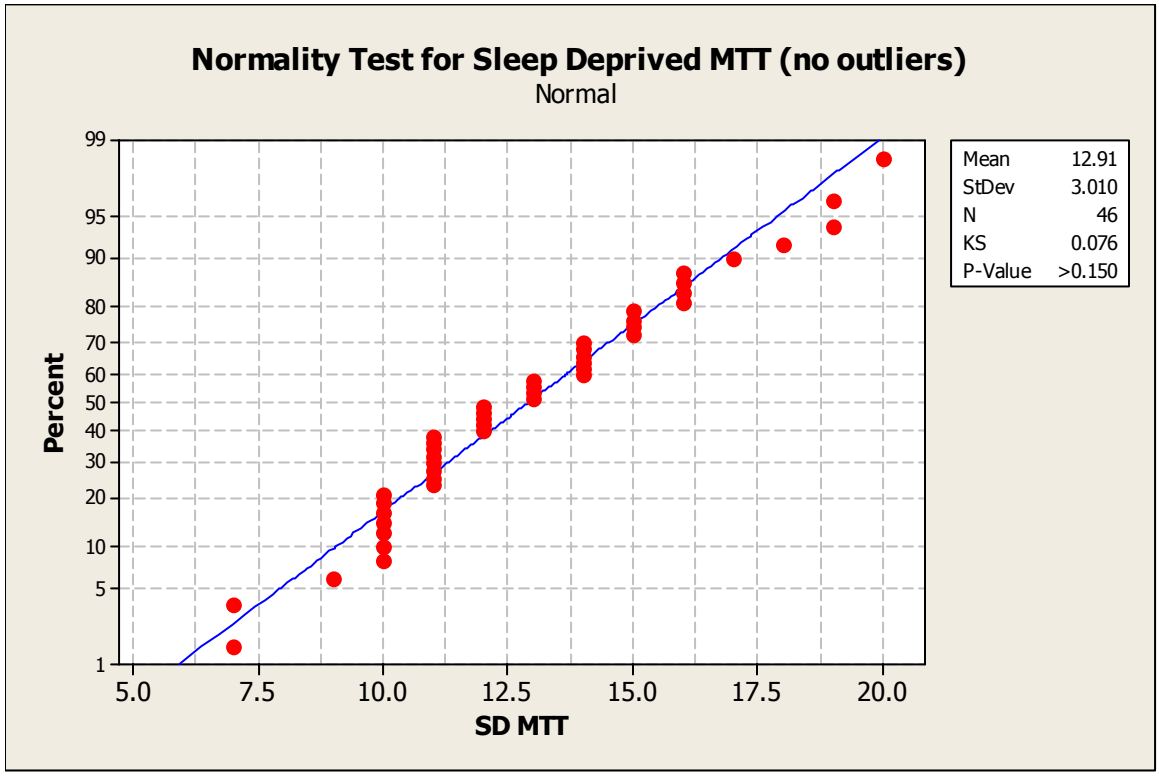


Figure D.9.5 Normality Test for Sleep-Deprived Maze Tracing Test with No Outliers

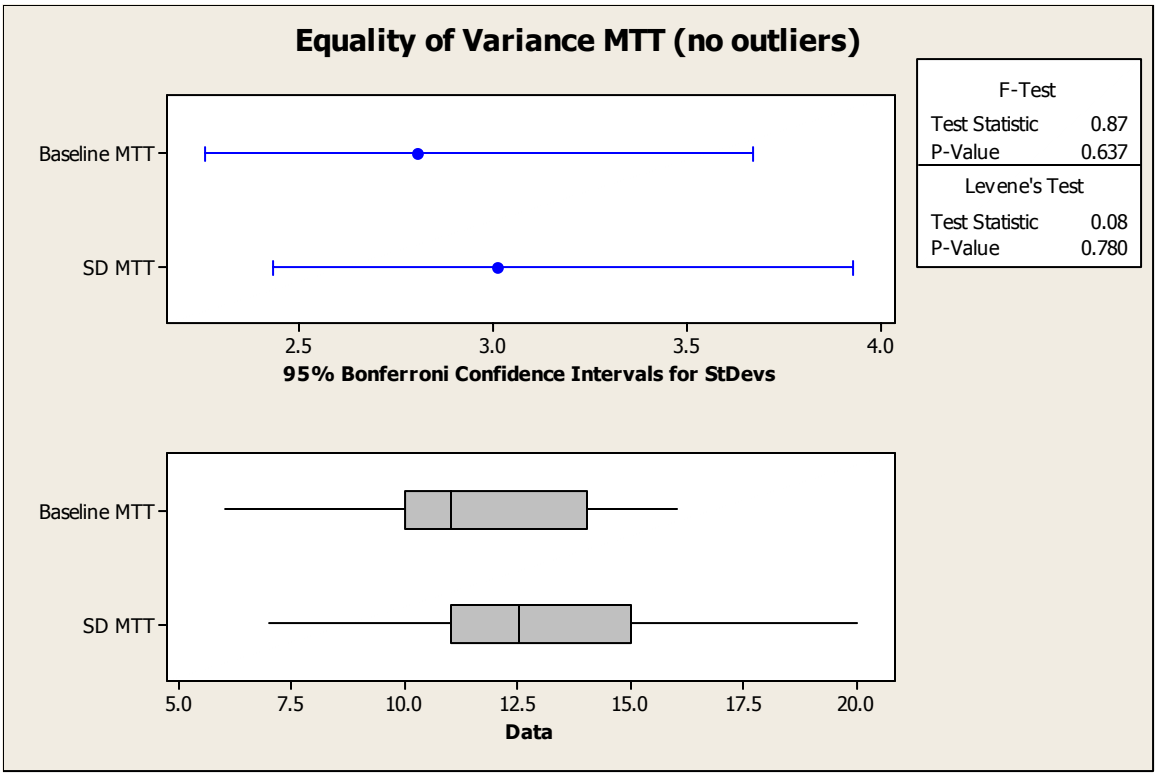


Figure D.9.6 Equality of Variance for Maze Tracing Test with No Outliers

D.10 MTT Improvement vs. Degradation

Table D.10.1 T-test for Baseline Scores (Improved vs. Degraded vs. No Difference)

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Improved*	*Degraded*
Mean	11.96875	12.8
Variance	10.93447581	6.177777778
Observations	32	10
Hypothesized Mean Difference	0	
df	20	
t Stat	-0.84862078	
P(T<=t) one-tail	0.203067603	
t Critical one-tail	1.724718218	
P(T<=t) two-tail	0.406135206	
t Critical two-tail	2.085963441	

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Improved*	*No Diff*
Mean	11.96875	11.16666667
Variance	10.93447581	8.566666667
Observations	32	6
Hypothesized Mean Difference	0	
df	8	
t Stat	0.602971305	
P(T<=t) one-tail	0.2816134	
t Critical one-tail	1.859548033	
P(T<=t) two-tail	0.5632268	
t Critical two-tail	2.306004133	

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Degraded*	*No Diff*
Mean	12.8	11.16666667
Variance	6.177777778	8.566666667
Observations	10	6
Hypothesized Mean Difference	0	
df	9	
t Stat	1.142008104	
P(T<=t) one-tail	0.141461234	
t Critical one-tail	1.833112923	
P(T<=t) two-tail	0.282922469	
t Critical two-tail	2.262157158	

D.10.1 Demographics and Improvement vs. Degradation

Table D.10.1.1 ANOVA for Improved vs. Degraded and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
age	Between Groups	16.233	2	8.117	.634	.535
	Within Groups	576.433	45	12.810		
	Total	592.667	47			
IQ	Between Groups	225.615	2	112.807	.760	.474
	Within Groups	6681.052	45	148.468		
	Total	6906.667	47			
education	Between Groups	5.233	2	2.617	.804	.454
	Within Groups	146.433	45	3.254		
	Total	151.667	47			
ME	Between Groups	3.965	2	1.982	.021	.979
	Within Groups	4285.952	45	95.243		
	Total	4289.917	47			

Table D.10.1.2 Chi Square Tests for Improved vs. Degraded by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	3 4.375 -1.375 6.25 50.00 8.57	1 0.625 0.375 2.08 16.67 20.00	2 1 1 4.17 33.33 25.00	6 12.50
1	26 23.333 2.6667 54.17 81.25 74.29	2 3.3333 -1.333 4.17 6.25 40.00	4 5.3333 -1.333 8.33 12.50 50.00	32 66.67
2	6 7.2917 -1.292 12.50 60.00 17.14	2 1.0417 0.9583 4.17 20.00 40.00	2 1.6667 0.3333 4.17 20.00 25.00	10 20.83
Total	35 72.92	5 10.42	8 16.67	48 100.00

Table D.10.1.3 Chi Square Tests for Improved vs. Degraded by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	4	4.0057	0.4052
Likelihood Ratio Chi-Square	4	3.7301	0.4438
Mantel-Haenszel Chi-Square	1	0.0607	0.8054
Phi Coefficient		0.2889	
Contingency Coefficient		0.2775	
Cramer's V		0.2043	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0607	0.8054
2	Row Mean Scores Differ	2	2.8799	0.2369
3	General Association	4	3.9223	0.4166

Total Sample Size = 48

Table D.10.1.4 Chi Square Tests for Improved vs. Degraded by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
0	3	3	6
	2.75	3.25	
	0.25	-0.25	
	6.25	6.25	12.50
	50.00	50.00	
	13.64	11.54	
1	15	17	32
	14.667	17.333	
	0.3333	-0.333	
	31.25	35.42	66.67
	46.88	53.13	
	68.18	65.38	
2	4	6	10
	4.5833	5.4167	
	-0.583	0.5833	
	8.33	12.50	20.83
	40.00	60.00	
	18.18	23.08	
Total	22	26	48
	45.83	54.17	100.00

Table D.10.1.5 Chi Square Tests for Improved vs. Degraded by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	2	0.1930	0.9080
Likelihood Ratio Chi-Square	2	0.1941	0.9075
Mantel-Haenszel Chi-Square	1	0.1748	0.6759
Phi Coefficient		0.0634	
Contingency Coefficient		0.0633	
Cramer's V		0.0634	

WARNING: 50% of the cells have expected counts less
than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1748	0.6759
2	Row Mean Scores Differ	2	0.1890	0.9098
3	General Association	2	0.1890	0.9098

Total Sample Size = 48

Table D.10.1.6 Chi Square Tests for Improved vs. Degraded by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	3	0	0	3	6
	3.125	0.875	0.25	1.75	
	-0.125	-0.875	-0.25	1.25	
	6.25	0.00	0.00	6.25	12.50
	50.00	0.00	0.00	50.00	
	12.00	0.00	0.00	21.43	
1	17	7	1	7	32
	16.667	4.6667	1.3333	9.3333	
	0.3333	2.3333	-0.333	-2.333	
	35.42	14.58	2.08	14.58	66.67
	53.13	21.88	3.13	21.88	
	68.00	100.00	50.00	50.00	
2	5	0	1	4	10
	5.2083	1.4583	0.4167	2.9167	
	-0.208	-1.458	0.5833	1.0833	
	10.42	0.00	2.08	8.33	20.83
	50.00	0.00	10.00	40.00	
	20.00	0.00	50.00	28.57	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.10.1.7 Chi Square Tests for Improved vs. Degraded by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	6	6.5486	0.3646
Likelihood Ratio Chi-Square	6	8.6055	0.1970
Mantel-Haenszel Chi-Square	1	0.0124	0.9114
Phi Coefficient		0.3694	
Contingency Coefficient		0.3465	
Cramer's V		0.2612	

WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0124	0.9114
2	Row Mean Scores Differ	2	1.5409	0.4628
3	General Association	6	6.4121	0.3786

Total Sample Size = 48

Table D.10.1.8 Chi Square Tests for Improved vs. Degraded by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	5	6	Total
0	0	5	0	0	1	6
	1.875	2.875	0.375	0.125	0.75	
	-1.875	2.125	-0.375	-0.125	0.25	
	0.00	10.42	0.00	0.00	2.08	12.50
	0.00	83.33	0.00	0.00	16.67	
	0.00	21.74	0.00	0.00	16.67	
1	14	14	2	1	1	32
	10	15.333	2	0.6667	4	
	4	-1.333	0	0.3333	-3	
	29.17	29.17	4.17	2.08	2.08	66.67
	43.75	43.75	6.25	3.13	3.13	
	93.33	60.87	66.67	100.00	16.67	
2	1	4	1	0	4	10
	3.125	4.7917	0.625	0.2083	1.25	
	-2.125	-0.792	0.375	-0.208	2.75	
	2.08	8.33	2.08	0.00	8.33	20.83
	10.00	40.00	10.00	0.00	40.00	
	6.67	17.39	33.33	0.00	66.67	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.10.1.9 Chi Square Tests for Improved vs. Degraded by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	8	16.2207	0.0393
Likelihood Ratio Chi-Square	8	17.5432	0.0249
Mantel-Haenszel Chi-Square	1	3.1677	0.0751
Phi Coefficient		0.5813	
Contingency Coefficient		0.5026	
Cramer's V		0.4111	

WARNING: 87% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	3.1677	0.0751
2	Row Mean Scores Differ	2	9.4180	0.0090
3	General Association	8	15.8828	0.0441

Total Sample Size = 48

Table D.10.1.10 Chi Square Tests for Improved vs. Degraded by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	1	5	0	6
	1.75	3.375	0.875	
	-0.75	1.625	-0.875	
	2.08	10.42	0.00	12.50
	16.67	83.33	0.00	
	7.14	18.52	0.00	
1	9	17	6	32
	9.3333	18	4.6667	
	-0.333	-1	1.3333	
	18.75	35.42	12.50	66.67
	28.13	53.13	18.75	
	64.29	62.96	85.71	
2	4	5	1	10
	2.9167	5.625	1.4583	
	1.0833	-0.625	-0.458	
	8.33	10.42	2.08	20.83
	40.00	50.00	10.00	
	28.57	18.52	14.29	
Total	14	27	7	48
	29.17	56.25	14.58	100.00

Table D.10.1.11 Chi Square Tests for Improved vs. Degraded by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	4	3.0431	0.5506
Likelihood Ratio Chi-Square	4	3.8234	0.4304
Mantel-Haenszel Chi-Square	1	0.3014	0.5830
Phi Coefficient		0.2518	
Contingency Coefficient		0.2442	
Cramer's V		0.1780	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.3014	0.5830
2	Row Mean Scores Differ	2	0.7694	0.6806
3	General Association	4	2.9797	0.5612

Total Sample Size = 48

D.11 Iowa Gambling Task Data (Normality and Equality of Variance)

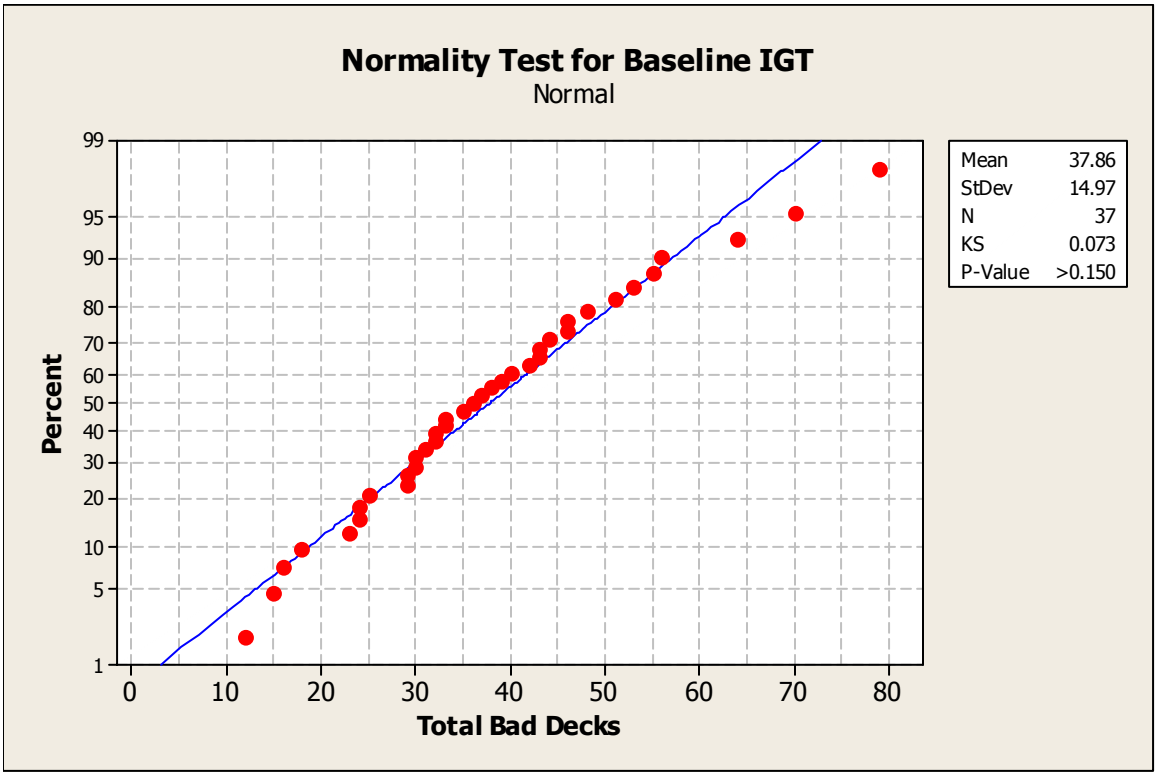


Figure D.11.1 Normality Test for Baseline Iowa Gambling Task

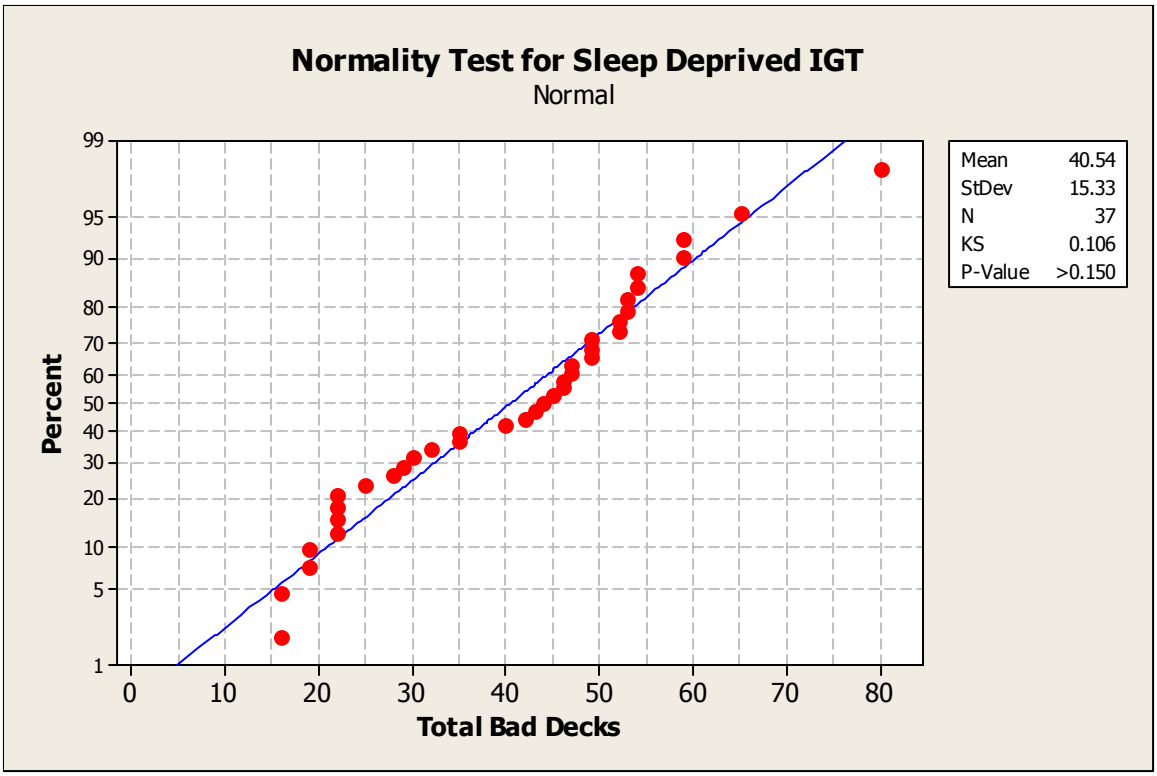


Figure D.11.2 Normality Test for Sleep Deprived Iowa Gambling Task

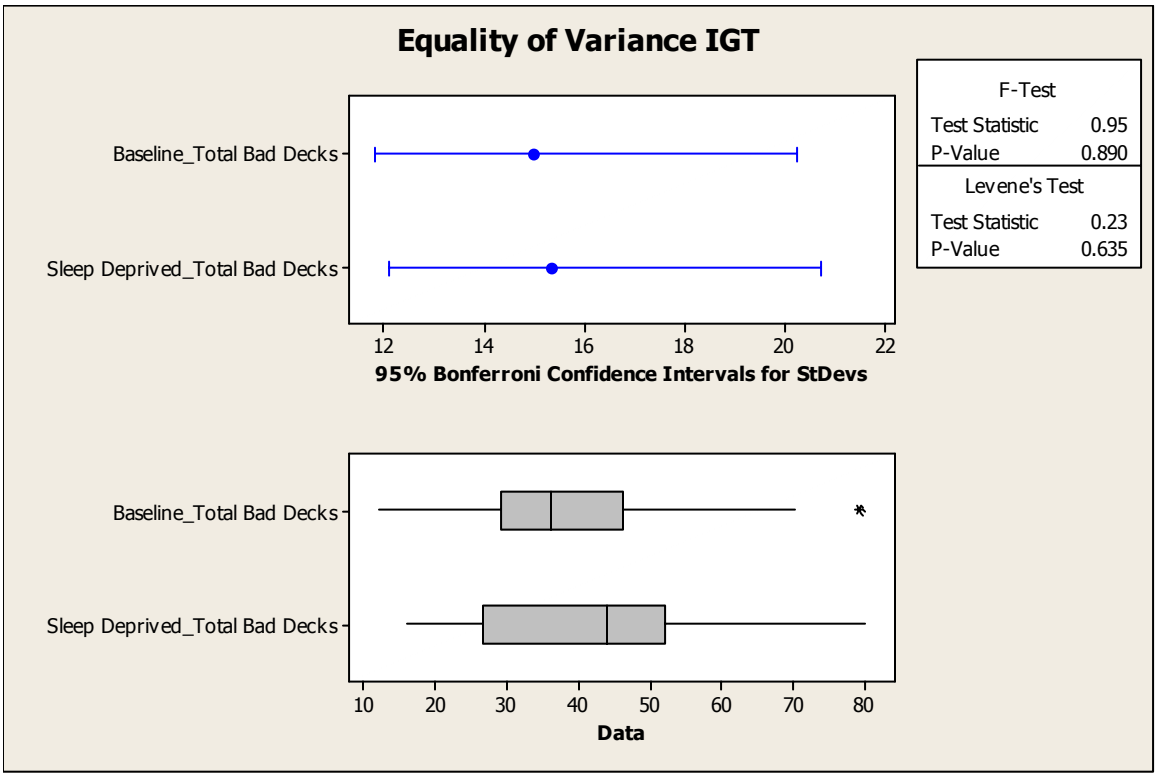


Figure D.11.3 Equality of Variance for Iowa Gambling Task

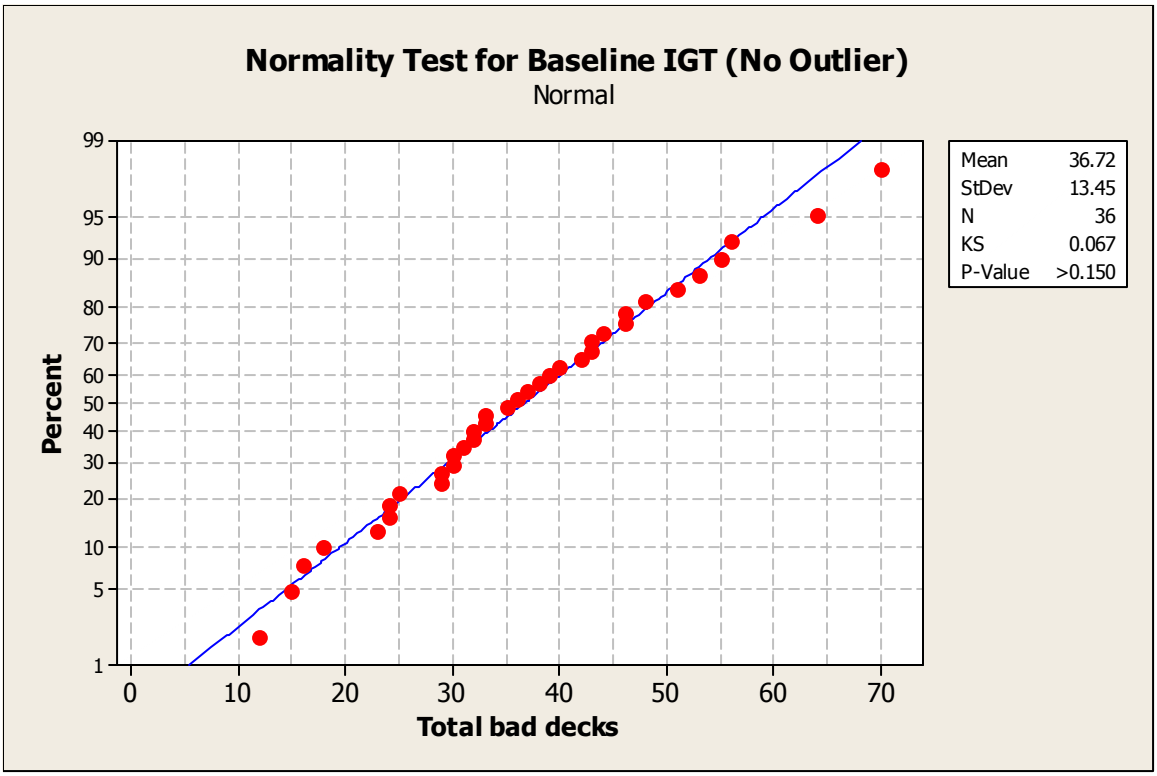


Figure D.11.4 Normality Test for Baseline Iowa Gambling Task with No Outliers

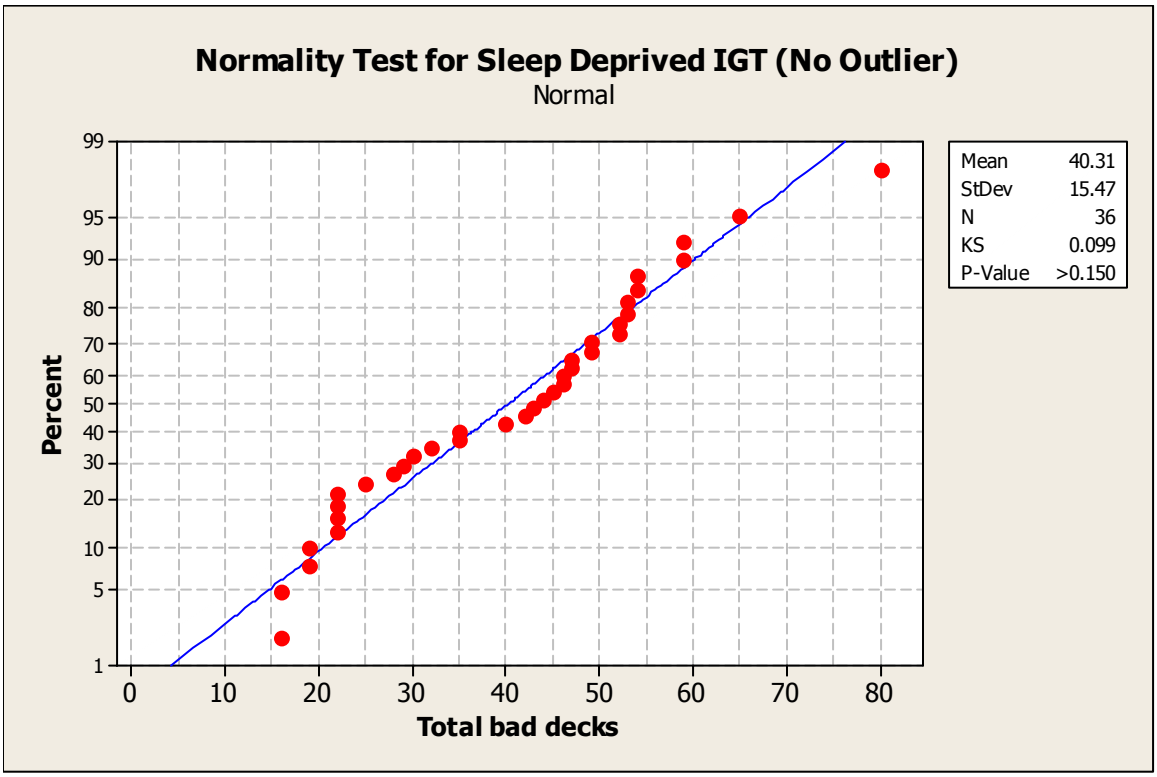


Figure D.11.5 Normality Test for Sleep Deprived Iowa Gambling Task with No Outliers

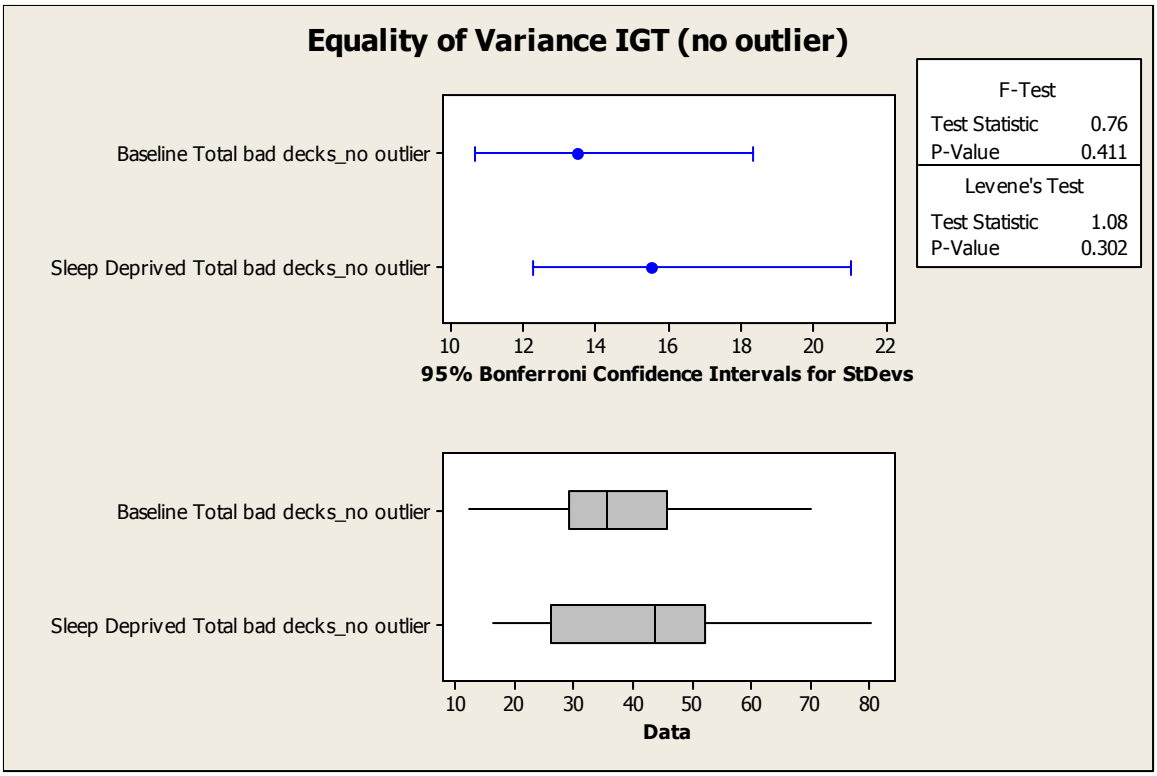


Figure D.11.6 Equality of Variance for Iowa Gambling Task with No Outliers

D.12 IGT Improvement vs. Degradation

Table D.12.1 T-test for Baselines Scores (Improved vs. Degraded vs. No Difference)

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Degraded*	*Improved*
Mean	34.52173913	46.18181818
Variance	118.5335968	367.7636364
Observations	23	11
Hypothesized Mean Difference	0	
df	13	
	-	
t Stat	1.877079721	
P(T<=t) one-tail	0.041564224	
t Critical one-tail	1.770933383	
P(T<=t) two-tail	0.083128448	
t Critical two-tail	2.160368652	

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Degraded*	*No Diff*
Mean	34.52173913	33
Variance	118.5335968	349
Observations	23	3
Hypothesized Mean Difference	0	
df	2	
t Stat	0.138062382	
P(T<=t) one-tail	0.451418533	
t Critical one-tail	2.91998558	
P(T<=t) two-tail	0.902837067	
t Critical two-tail	4.30265273	

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Improved*	*No Diff*
Mean	46.18181818	33
Variance	367.7636364	349
Observations	11	3
Hypothesized Mean Difference	0	
df	3	
t Stat	1.077130031	
P(T<=t) one-tail	0.180161246	
t Critical one-tail	2.353363435	
P(T<=t) two-tail	0.360322492	
t Critical two-tail	3.182446305	

D.12.1 Demographics and Improvement vs. Degradation

Table D.12.1.1 ANOVA for Improved vs. Degraded and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
age	Between Groups	11.987	2	5.994	.409	.668
	Within Groups	498.445	34	14.660		
	Total	510.432	36			
IQ	Between Groups	1063.219	2	531.609	3.900	.030
	Within Groups	4635.051	34	136.325		
	Total	5698.270	36			
education	Between Groups	2.772	2	1.386	.484	.620
	Within Groups	97.336	34	2.863		
	Total	100.108	36			
ME	Between Groups	164.186	2	82.093	.852	.436
	Within Groups	3276.516	34	96.368		
	Total	3440.703	36			

D.12.1.1 Post Hoc WASI Analysis

Table D.12.1.1.1 T-tests for Improved vs. Degraded based on WASI Score

	Imp/Deg	N	Mean	Std. Deviation	Std. Error Mean
IQ	Improve	11	109.7273	9.66531	2.91420
	Degrade	23	107.6957	12.79714	2.66839

Table D.12.1.1.2 T-tests for Improved vs. Degraded based on WASI Score Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IQ	Equal variances assumed	.826	.370	.465	32	.645	2.03162	4.36507	-6.8597	10.9227
	Equal variances not assumed			.514	25.614	.612	2.03162	3.95131	-6.0968	10.1596

Table D.12.1.1.3 T-tests for Improved vs. No Difference based on WASI Score Group Statistics

	Imp/Deg	N	Mean	Std. Deviation	Std. Error Mean
IQ	Improve	11	109.7273	9.66531	2.91420
	No Diff	3	89.0000	7.00000	4.04145

Table D.12.1.1.4 T-tests for Improved vs. No Difference based on WASI Score Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IQ	Equal variances assumed	.388	.545	3.431	12	.005	20.7272	6.04080	7.56550	33.8890
	Equal variances not assumed			4.160	4.383	.012	20.7272	4.98256	7.35795	34.0965

Table D.12.1.1.5 T-tests for No Difference vs. Degraded based on WASI Score Group Statistics

	Imp/Deg	N	Mean	Std. Deviation	Std. Error Mean
IQ	No Diff	3	89.0000	7.00000	4.04145
	Degrade	23	107.6957	12.79714	2.66839

Table D.12.1.1.6 T-tests for No Difference vs. Degraded based on WASI Score Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IQ	Equal variances assumed	.963	.336	-2.453	24	.022	-18.6956	7.62269	-34.428	-2.9631
	Equal variances not assumed			-3.860	4.054	.018	-18.6956	4.84289	-32.071	-5.3197

Table D.12.1.1.7 WASI ANOVA with Language and Education Covariates

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1159.019(a)	4	289.755	2.043	.112
Intercept	3365.175	1	3365.175	23.723	.000
Education	83.953	1	83.953	.592	.447
Lang	.031	1	.031	.000	.988
imp1deg2	1112.883	2	556.442	3.923	.030
Error	4539.251	32	141.852		
Total	427601.000	37			
Corrected Total	5698.270	36			

a R Squared = .203 (Adjusted R Squared = .104)

Table D.12.1.1.8 WASI ANOVA with Education Covariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1158.988(a)	3	386.329	2.809	.055
Intercept	3449.756	1	3449.756	25.079	.000
education	95.769	1	95.769	.696	.410
imp1deg2	1118.589	2	559.295	4.066	.026
Error	4539.282	33	137.554		
Total	427601.000	37			
Corrected Total	5698.270	36			

a R Squared = .203 (Adjusted R Squared = .131)

Table D.12.1.1.9 WASI ANOVA with Language Covariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1075.066(a)	3	358.355	2.558	.072
Intercept	64027.655	1	64027.655	457.023	.000
lang	11.847	1	11.847	.085	.773
imp1deg2	1075.015	2	537.508	3.837	.032
Error	4623.204	33	140.097		
Total	427601.000	37			
Corrected Total	5698.270	36			

a R Squared = .189 (Adjusted R Squared = .115)

Table D.12.1.2 Chi Square Test for Improved vs. Degraded by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	2 2.3514 -0.351 5.41 66.67 6.90	0 0.1622 -0.162 0.00 0.00 0.00	1 0.4865 0.5135 2.70 33.33 16.67	3 8.11
1	8 8.6216 -0.622 21.62 72.73 27.59	1 0.5946 0.4054 2.70 9.09 50.00	2 1.7838 0.2162 5.41 18.18 33.33	11 29.73
2	19 18.027 0.973 51.35 82.61 65.52	1 1.2432 -0.243 2.70 4.35 50.00	3 3.7297 -0.73 8.11 13.04 50.00	23 62.16
Total	29 78.38	2 5.41	6 16.22	37 100.00

Table D.12.1.3 Chi Square Test for Improved vs. Degraded by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	4	1.3470	0.8533
Likelihood Ratio Chi-Square	4	1.3496	0.8529
Mantel-Haenszel Chi-Square	1	0.7547	0.3850
Phi Coefficient		0.1908	
Contingency Coefficient		0.1874	
Cramer's V		0.1349	

WARNING: 78% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.7547	0.3850
2	Row Mean Scores Differ	2	0.7637	0.6826
3	General Association	4	1.3106	0.8596

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Table D.12.1.4 Chi Square Test for Improved vs. Degraded by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
0	2	1	3
	1.2973	1.7027	
	0.7027	-0.703	
	5.41	2.70	8.11
	66.67	33.33	
	12.50	4.76	
1	4	7	11
	4.7568	6.2432	
	-0.757	0.7568	
	10.81	18.92	29.73
	36.36	63.64	
	25.00	33.33	
2	10	13	23
	9.9459	13.054	
	0.0541	-0.054	
	27.03	35.14	62.16
	43.48	56.52	
	62.50	61.90	
Total	16	21	37
	43.24	56.76	100.00

Frequency Missing = 11

Table D.12.1.5 Chi Square Test for Improved vs. Degraded by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	2	0.8833	0.6430
Likelihood Ratio Chi-Square	2	0.8831	0.6430
Mantel-Haenszel Chi-Square	1	0.1098	0.7404
Phi Coefficient		0.1545	
Contingency Coefficient		0.1527	
Cramer's V		0.1545	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1098	0.7404
2	Row Mean Scores Differ	2	0.8594	0.6507
3	General Association	2	0.8594	0.6507

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Table D.12.1.6 Chi Square Test for Improved vs. Degraded by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
0	2	1	0	0	3
	1.5405	0.4054	0.1622	0.8919	
	0.4595	0.5946	-0.162	-0.892	
	5.41	2.70	0.00	0.00	8.11
	66.67	33.33	0.00	0.00	
	10.53	20.00	0.00	0.00	
1	6	3	0	2	11
	5.6486	1.4865	0.5946	3.2703	
	0.3514	1.5135	-0.595	-1.27	
	16.22	8.11	0.00	5.41	29.73
	54.55	27.27	0.00	18.18	
	31.58	60.00	0.00	18.18	
2	11	1	2	9	23
	11.811	3.1081	1.2432	6.8378	
	-0.811	-2.108	0.7568	2.1622	
	29.73	2.70	5.41	24.32	62.16
	47.83	4.35	8.70	39.13	
	57.89	20.00	100.00	81.82	
Total	19	5	2	11	37
	51.35	13.51	5.41	29.73	100.00

Frequency Missing = 11

Table D.12.1.7 Chi Square Test for Improved vs. Degraded by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	6	7.3439	0.2902
Likelihood Ratio Chi-Square	6	8.8348	0.1831
Mantel-Haenszel Chi-Square	1	2.5367	0.1112
Phi Coefficient		0.4455	
Contingency Coefficient		0.4070	
Cramer's V		0.3150	

WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.5367	0.1112
2	Row Mean Scores Differ	2	2.5426	0.2805
3	General Association	6	7.1454	0.3076

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Table D.12.1.8 Chi Square Test for Improved vs. Degraded by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	6	Total
0	0	2	0	1	3
	0.8108	1.6216	0.2432	0.3243	
	-0.811	0.3784	-0.243	0.6757	
	0.00	5.41	0.00	2.70	8.11
	0.00	66.67	0.00	33.33	
	0.00	10.00	0.00	25.00	
1	5	4	1	1	11
	2.973	5.9459	0.8919	1.1892	
	2.027	-1.946	0.1081	-0.189	
	13.51	10.81	2.70	2.70	29.73
	45.45	36.36	9.09	9.09	
	50.00	20.00	33.33	25.00	
2	5	14	2	2	23
	6.2162	12.432	1.8649	2.4865	
	-1.216	1.5676	0.1351	-0.486	
	13.51	37.84	5.41	5.41	62.16
	21.74	60.87	8.70	8.70	
	50.00	70.00	66.67	50.00	
Total	10	20	3	4	37
	27.03	54.05	8.11	10.81	100.00

Table D.12.1.9 Chi Square Test for Improved vs. Degraded by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	6	5.1527	0.5244
Likelihood Ratio Chi-Square	6	5.5573	0.4746
Mantel-Haenszel Chi-Square	1	0.3509	0.5536
Phi Coefficient		0.3732	
Contingency Coefficient		0.3496	
Cramer's V		0.2639	

WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.3509	0.5536
2	Row Mean Scores Differ	2	1.6059	0.4480
3	General Association	6	5.0134	0.5421

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Table D.12.1.10 Chi Square Test for Improved vs. Degraded by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
0	0 0.7297 -0.73 0.00 0.00 0.00	3 1.8649 1.1351 8.11 100.00 13.04	0 0.4054 -0.405 0.00 0.00 0.00	3 8.11
1	3 2.6757 0.3243 8.11 27.27 33.33	7 6.8378 0.1622 18.92 63.64 30.43	1 1.4865 -0.486 2.70 9.09 20.00	11 29.73
2	6 5.5946 0.4054 16.22 26.09 66.67	13 14.297 -1.297 35.14 56.52 56.52	4 3.1081 0.8919 10.81 17.39 80.00	23 62.16
Total	9 24.32	23 62.16	5 13.51	37 100.00

Frequency Missing = 11

Table D.12.1.11 Chi Square Test for Improved vs. Degraded by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	4	2.4315	0.6569
Likelihood Ratio Chi-Square	4	3.4589	0.4842
Mantel-Haenszel Chi-Square	1	0.0046	0.9460
Phi Coefficient		0.2564	
Contingency Coefficient		0.2483	
Cramer's V		0.1813	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

Summary Statistics for imp1deg2 by MEtype

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0046	0.9460
2	Row Mean Scores Differ	2	0.2789	0.8698
3	General Association	4	2.3658	0.6688

Effective Sample Size = 37
Frequency Missing = 11

WARNING: 23% of the data are missing.

D.13 Psychomotor Vigilance Task (PVT) Data (Normality and Equality of Variance)

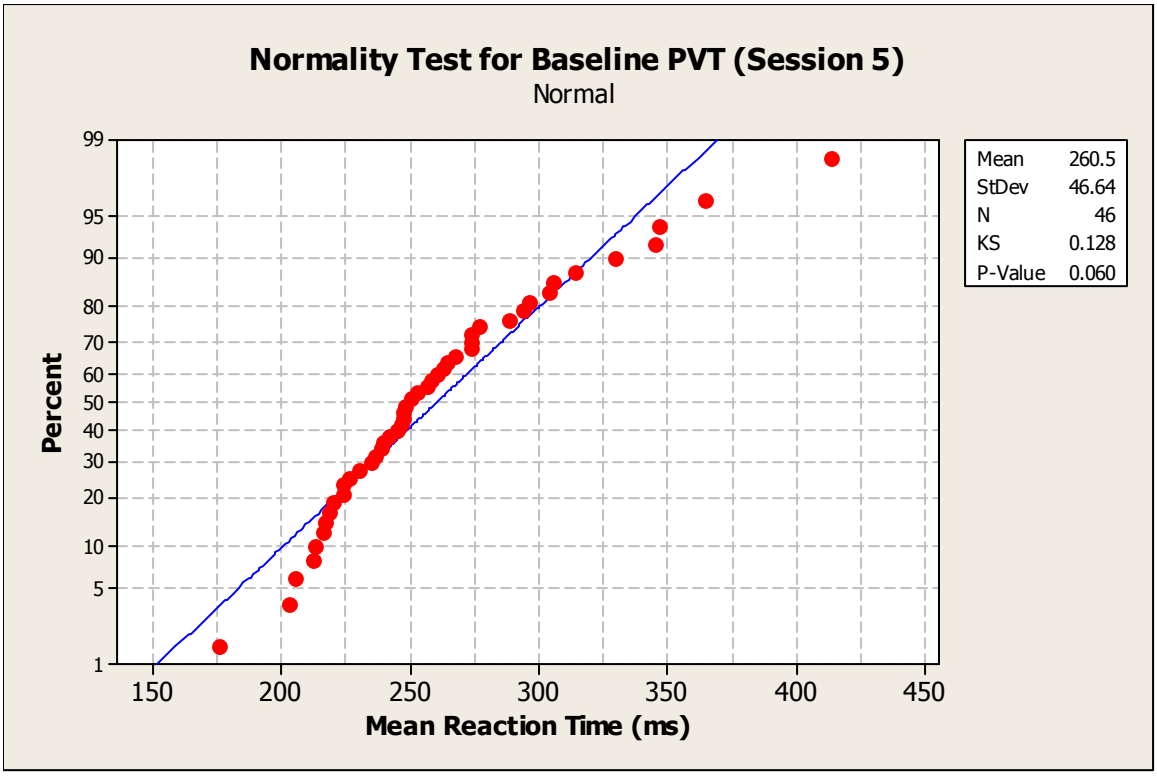


Figure D.13.1 Normality test for PVT Session 5 (Baseline) data

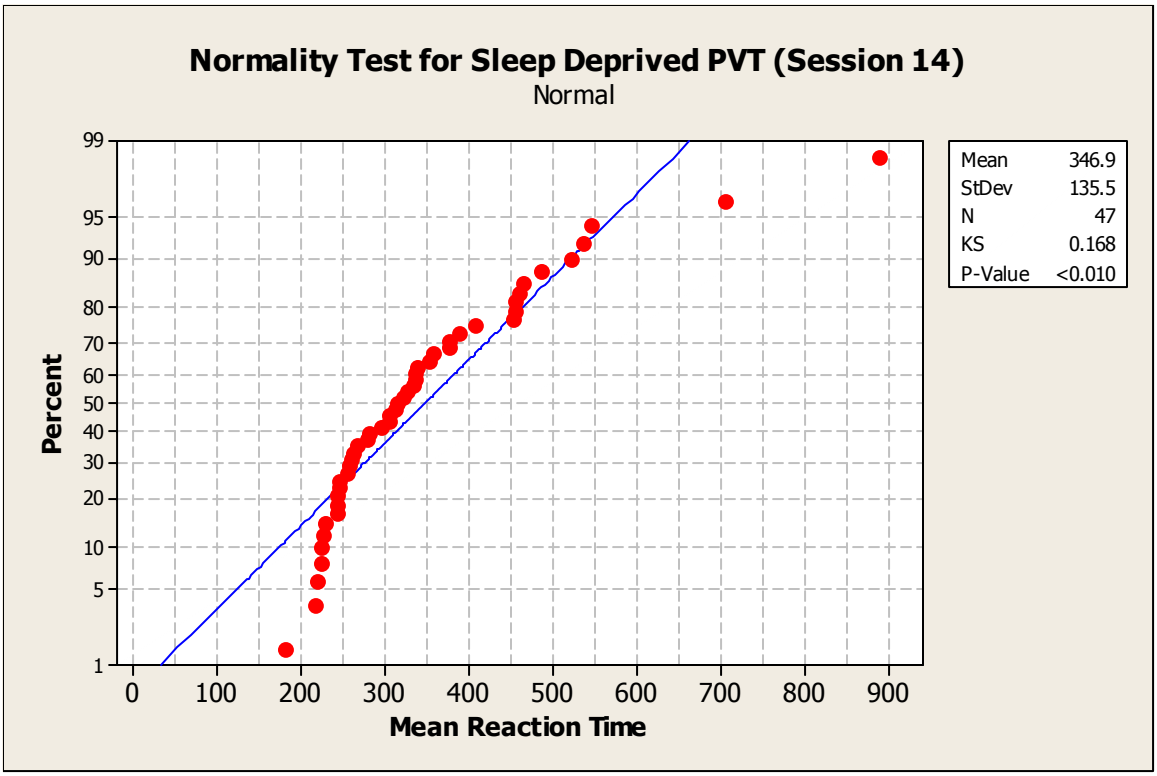


Figure D.13.2 Normality test for PVT Session 14 (sleep deprived) data

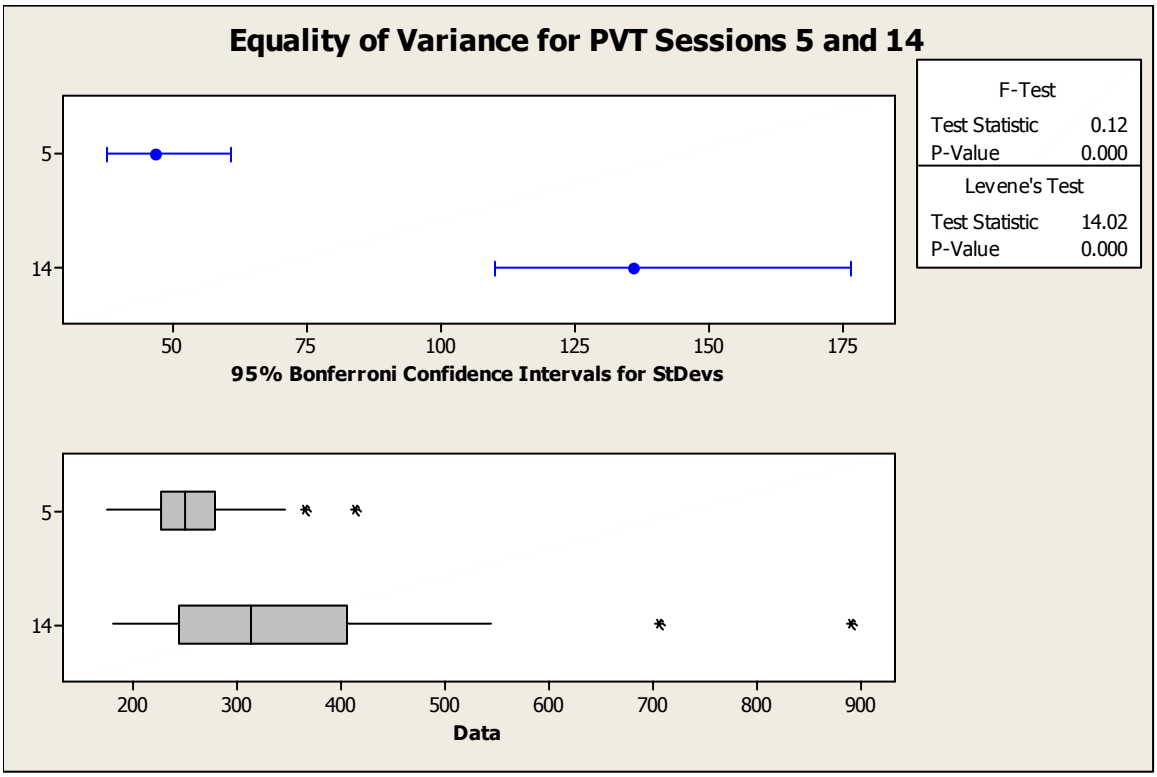


Figure D.13.3 Test of Equal Variance for PVT

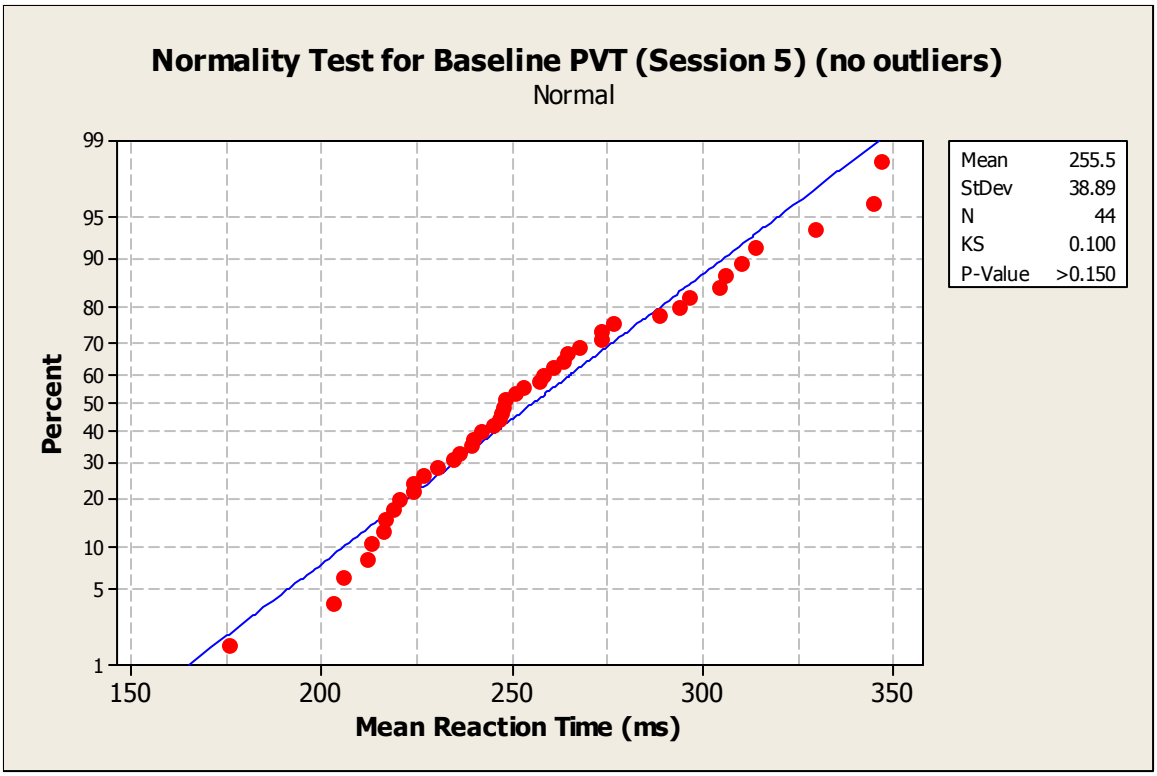


Figure D.13.4 Normality Test for Baseline PVT (no outliers)

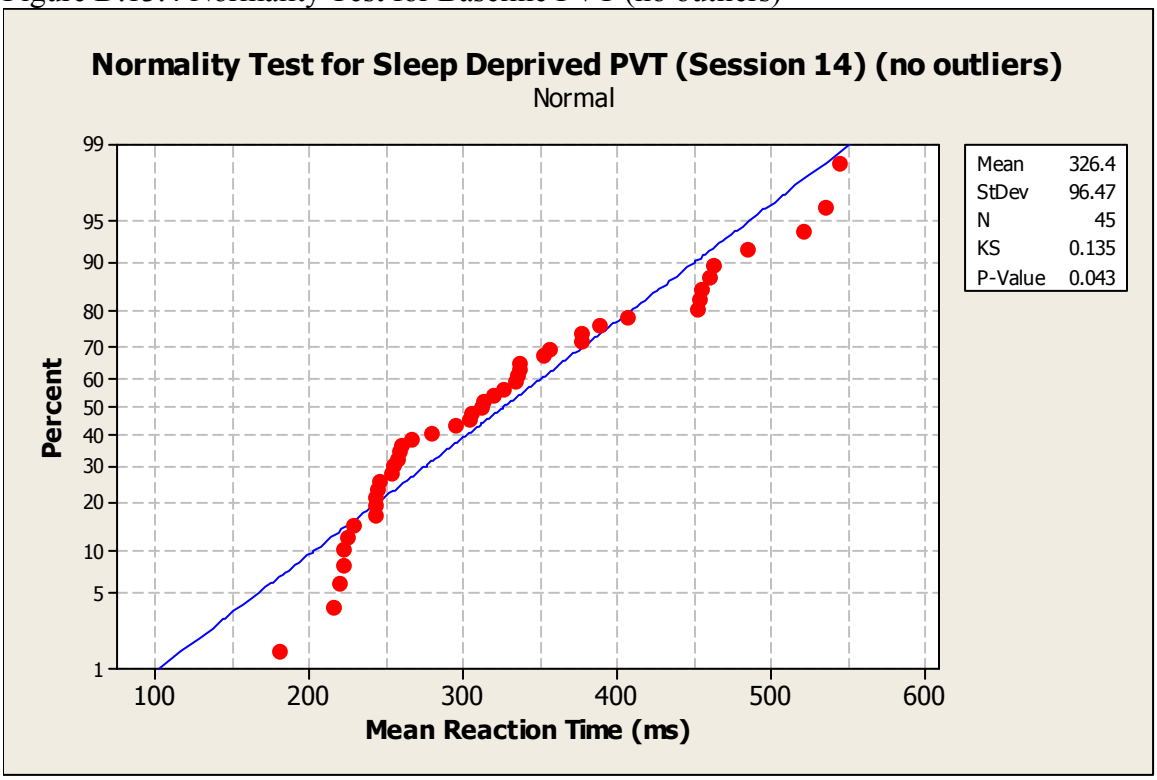


Figure D.13.5 Normality Test for Sleep Deprived PVT (no outliers)

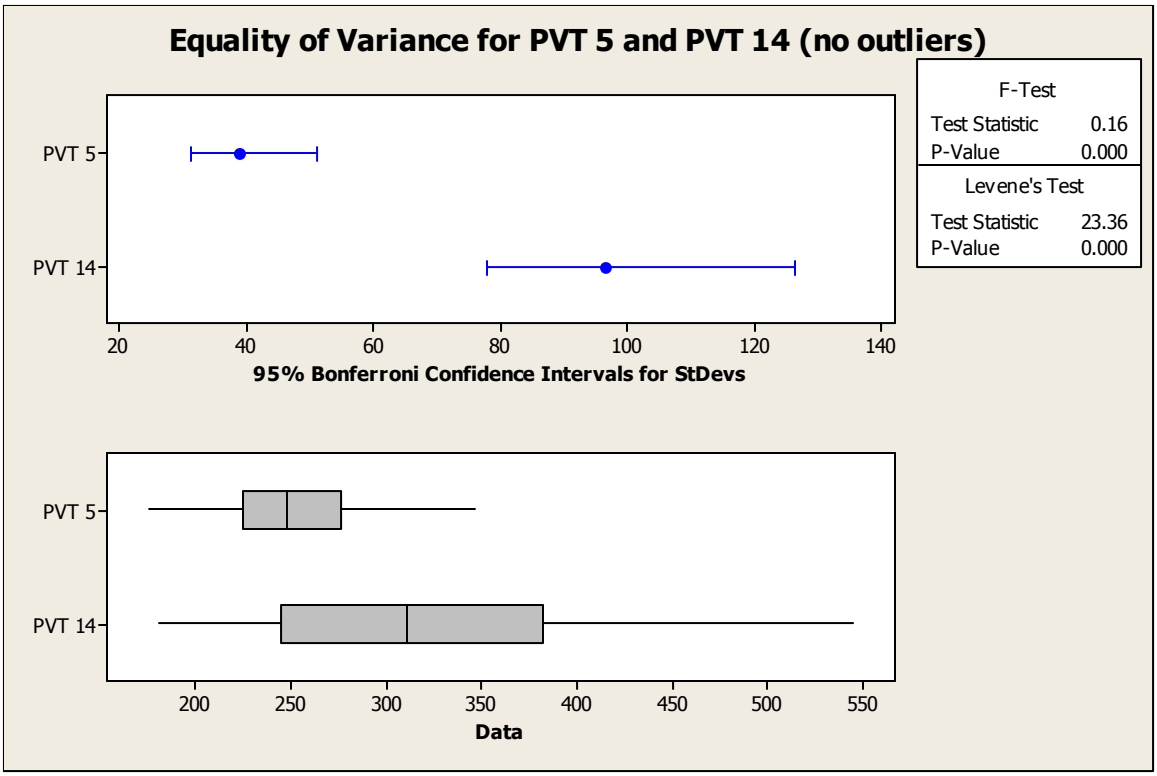


Figure D.13.6 Equality of Variance for PVT (no outliers)

D.14 PVT Improvement vs. Degradation

Table D.14.1 T-test for Baseline Scores (Improved vs. Degraded)

t-Test: Two-Sample Assuming Unequal Variances

MEAN BASELINE SCORES	*Degraded*	*Improved*
Mean	262.252	248.5266667
Variance	2456.91875	219.8114667
Observations	40	6
Hypothesized Mean Difference	0	
Df	26	
t Stat	1.386056433	
P(T<=t) one-tail	0.088751619	
t Critical one-tail	1.705617901	
P(T<=t) two-tail	0.177503239	
t Critical two-tail	2.055529418	

D.14.1 Demographics and Improvement vs. Degradation

Table D.14.1 ANOVA for Improved vs. Degraded and Demographics

		Sum of Squares	df	Mean Square	F	Sig.
age	Between Groups	.074	1	.074	.006	.940
	Within Groups	592.592	46	12.882		
	Total	592.667	47			
IQ	Between Groups	52.323	1	52.323	.342	.562
	Within Groups	7046.990	46	153.195		
	Total	7099.313	47			
education	Between Groups	.196	1	.196	.060	.808
	Within Groups	151.470	46	3.293		
	Total	151.667	47			
ME	Between Groups	20.468	1	20.468	.221	.641
	Within Groups	4261.345	46	92.638		
	Total	4281.813	47			

Table D.14.1.2 Chi-Square Tests for Improved vs. Degraded by Language Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	Total
1	5 4.9583 0.0417 10.42 71.43 14.71	0 0.7292 -0.729 0.00 0.00 0.00	2 1.3125 0.6875 4.17 28.57 22.22	7 14.58
2	29 29.042 -0.042 60.42 70.73 85.29	5 4.2708 0.7292 10.42 12.20 100.00	7 7.6875 -0.688 14.58 17.07 77.78	41 85.42
Total	34 70.83	5 10.42	9 18.75	48 100.00

Table D.14.1.3 Chi-Square Tests for Improved vs. Degraded by Language

Statistics for Table of imp1deg2 by language

Statistic	DF	Value	Prob
Chi-Square	2	1.2757	0.5284
Likelihood Ratio Chi-Square	2	1.9500	0.3772
Mantel-Haenszel Chi-Square	1	0.1094	0.7409
Phi Coefficient		0.1630	
Contingency Coefficient		0.1609	
Cramer's V		0.1630	

WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by language

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1094	0.7409
2	Row Mean Scores Differ	1	0.1094	0.7409
3	General Association	2	1.2491	0.5355

Total Sample Size = 48

Table D.14.1.4 Chi-Square Tests for Improved vs. Degraded by Gender Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	Total
1	4	3	7
	3.0625	3.9375	
	0.9375	-0.938	
	8.33	6.25	14.58
	57.14	42.86	
	19.05	11.11	
2	17	24	41
	17.938	23.063	
	-0.938	0.9375	
	35.42	50.00	85.42
	41.46	58.54	
	80.95	88.89	
Total	21	27	48
	43.75	56.25	100.00

Table D.14.1.5 Chi-Square Tests for Improved vs. Degraded by Gender

Statistics for Table of imp1deg2 by gender

Statistic	DF	Value	Prob
Chi-Square	1	0.5973	0.4396
Likelihood Ratio Chi-Square	1	0.5924	0.4415
Continuity Adj. Chi-Square	1	0.1301	0.7183
Mantel-Haenszel Chi-Square	1	0.5849	0.4444
Phi Coefficient		0.1116	
Contingency Coefficient		0.1109	
Cramer's V		0.1116	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Cell (1,1) Frequency (F)	4
Left-sided Pr <= F	0.8815
Right-sided Pr >= F	0.3562
Table Probability (P)	0.2378
Two-sided Pr <= P	0.6830

Sample Size = 48

Summary Statistics for imp1deg2 by gender

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.5849	0.4444
2	Row Mean Scores Differ	1	0.5849	0.4444
3	General Association	1	0.5849	0.4444

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	1.8824	0.3722	9.5190
	Logit	1.8824	0.3722	9.5190
Cohort (Col1 Risk)	Mantel-Haenszel	1.3782	0.6592	2.8812
	Logit	1.3782	0.6592	2.8812
Cohort (Col2 Risk)	Mantel-Haenszel	0.7321	0.2997	1.7888
	Logit	0.7321	0.2997	1.7888

Total Sample Size = 48

Table D.14.1.6 Chi-Square Tests for Improved vs. Degraded by Occupation Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
1	3	2	0	2	7
	3.6458	1.0208	0.2917	2.0417	
	-0.646	0.9792	-0.292	-0.042	
	6.25	4.17	0.00	4.17	14.58
	42.86	28.57	0.00	28.57	
	12.00	28.57	0.00	14.29	
2	22	5	2	12	41
	21.354	5.9792	1.7083	11.958	
	0.6458	-0.979	0.2917	0.0417	
	45.83	10.42	4.17	25.00	85.42
	53.66	12.20	4.88	29.27	
	88.00	71.43	100.00	85.71	
Total	25	7	2	14	48
	52.08	14.58	4.17	29.17	100.00

Table D.14.1.7 Chi-Square Tests for Improved vs. Degraded by Occupation

Statistics for Table of imp1deg2 by occupation

Statistic	DF	Value	Prob
Chi-Square	3	1.5759	0.6649
Likelihood Ratio Chi-Square	3	1.6744	0.6426
Mantel-Haenszel Chi-Square	1	0.0070	0.9334
Phi Coefficient		0.1812	
Contingency Coefficient		0.1783	
Cramer's V		0.1812	

WARNING: 63% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by occupation

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0070	0.9334
2	Row Mean Scores Differ	1	0.0070	0.9334
3	General Association	3	1.5431	0.6724

Total Sample Size = 48

Table D.14.1.8 Chi-Square Tests for Improved vs. Degraded by Ethnicity Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	4	5	6	Total
1	2	4	0	0	1	7
	2.1875	3.3542	0.4375	0.1458	0.875	
	-0.188	0.6458	-0.438	-0.146	0.125	
	4.17	8.33	0.00	0.00	2.08	14.58
	28.57	57.14	0.00	0.00	14.29	
	13.33	17.39	0.00	0.00	16.67	
2	13	19	3	1	5	41
	12.813	19.646	2.5625	0.8542	5.125	
	0.1875	-0.646	0.4375	0.1458	-0.125	
	27.08	39.58	6.25	2.08	10.42	85.42
	31.71	46.34	7.32	2.44	12.20	
	86.67	82.61	100.00	100.00	83.33	
Total	15	23	3	1	6	48
	31.25	47.92	6.25	2.08	12.50	100.00

Table D.14.1.9 Chi-Square Tests for Improved vs. Degraded by Ethnicity

Statistics for Table of imp1deg2 by ethnicity

Statistic	DF	Value	Prob
Chi-Square	4	0.8682	0.9291
Likelihood Ratio Chi-Square	4	1.4390	0.8374
Mantel-Haenszel Chi-Square	1	0.0245	0.8756
Phi Coefficient		0.1345	
Contingency Coefficient		0.1333	
Cramer's V		0.1345	

WARNING: 70% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by ethnicity

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0245	0.8756
2	Row Mean Scores Differ	1	0.0245	0.8756
3	General Association	4	0.8501	0.9316

Total Sample Size = 48

Table D.14.1.10 Chi-Square Tests for Improved vs. Degraded by Morningness-Eveningness Cross Tabulation Matrix

Frequency Expected Deviation Percent Row Pct Col Pct	1	2	3	4	Total
1	2	3	2	0	7
	2.9167	2.1875	1.75	0.1458	
	-0.917	0.8125	0.25	-0.146	
	4.17	6.25	4.17	0.00	14.58
	28.57	42.86	28.57	0.00	
	10.00	20.00	16.67	0.00	
2	18	12	10	1	41
	17.083	12.813	10.25	0.8542	
	0.9167	-0.813	-0.25	0.1458	
	37.50	25.00	20.83	2.08	85.42
	43.90	29.27	24.39	2.44	
	90.00	80.00	83.33	100.00	
Total	20	15	12	1	48
	41.67	31.25	25.00	2.08	100.00

Table D.14.1.11 Chi-Square Tests for Improved vs. Degraded by Morningness-Eveningness

Statistics for Table of imp1deg2 by MEtype

Statistic	DF	Value	Prob
Chi-Square	3	0.9031	0.8247
Likelihood Ratio Chi-Square	3	1.0508	0.7890
Mantel-Haenszel Chi-Square	1	0.1707	0.6795
Phi Coefficient		0.1372	
Contingency Coefficient		0.1359	
Cramer's V		0.1372	

WARNING: 63% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 48

Summary Statistics for imp1deg2 by MEtype
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.1707	0.6795
2	Row Mean Scores Differ	1	0.1707	0.6795
3	General Association	3	0.8843	0.8292

Total Sample Size = 48

D.15 PVT Log Transform Data (Normality and Equality of Variance)

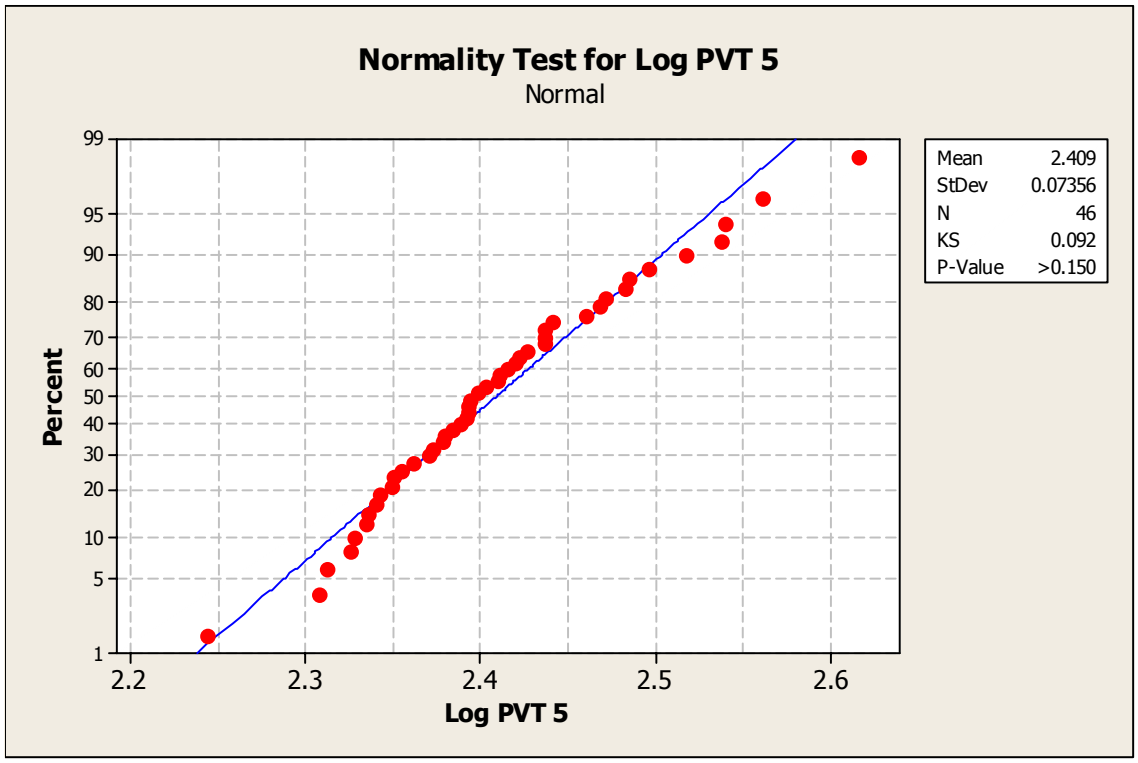


Figure D.15.1 Normality Test for Baseline PVT Log Transform

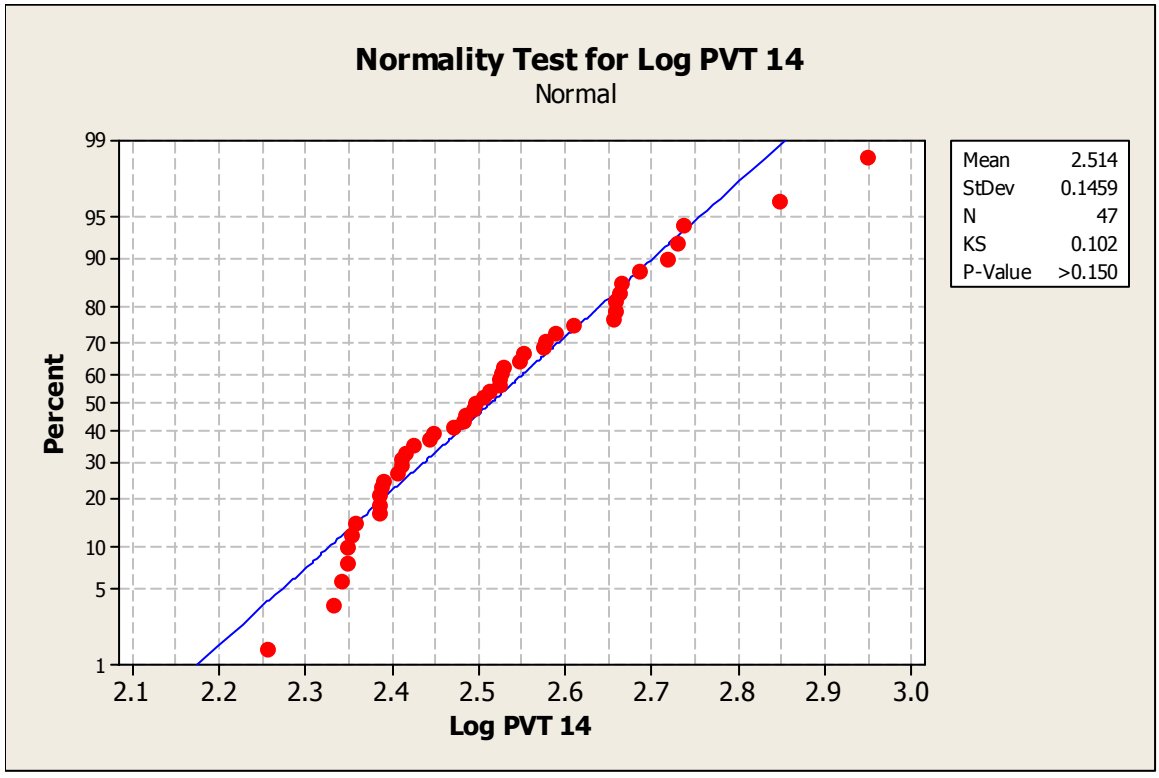


Figure D.15.2 Normality Test for Sleep Deprived PVT Log Transform

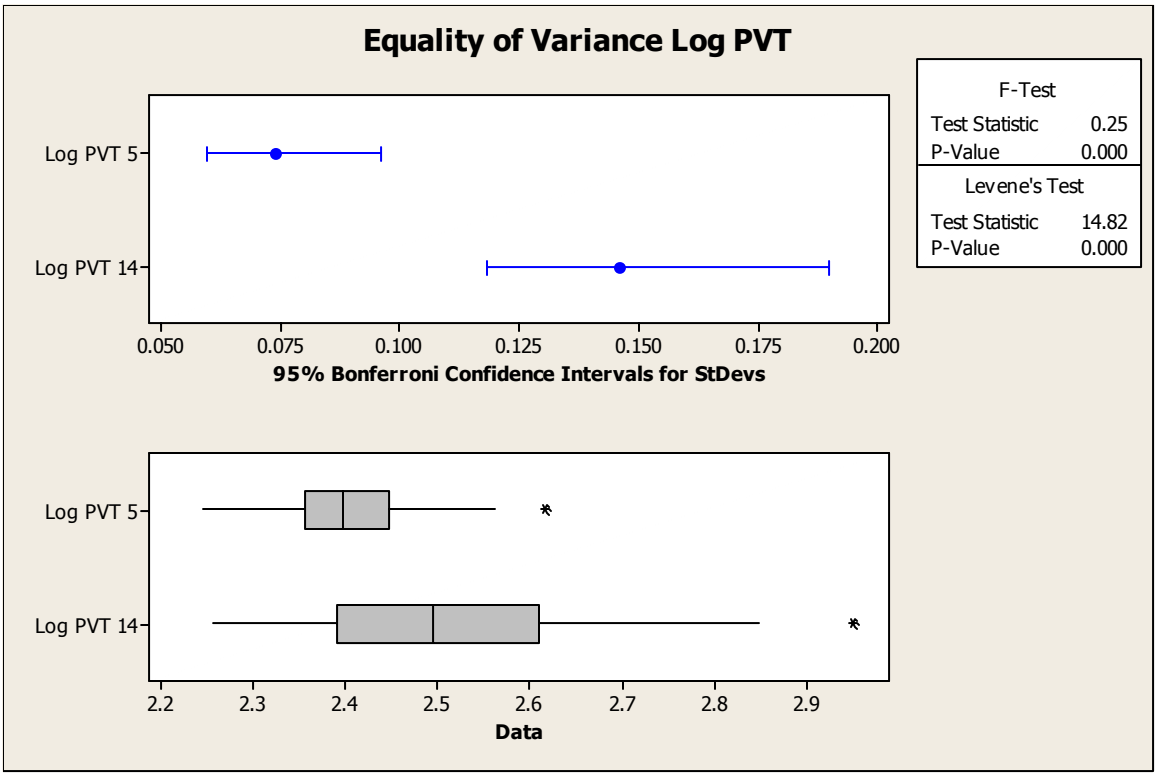


Figure D.15.3 Equality of Variance for PVT Log Transform

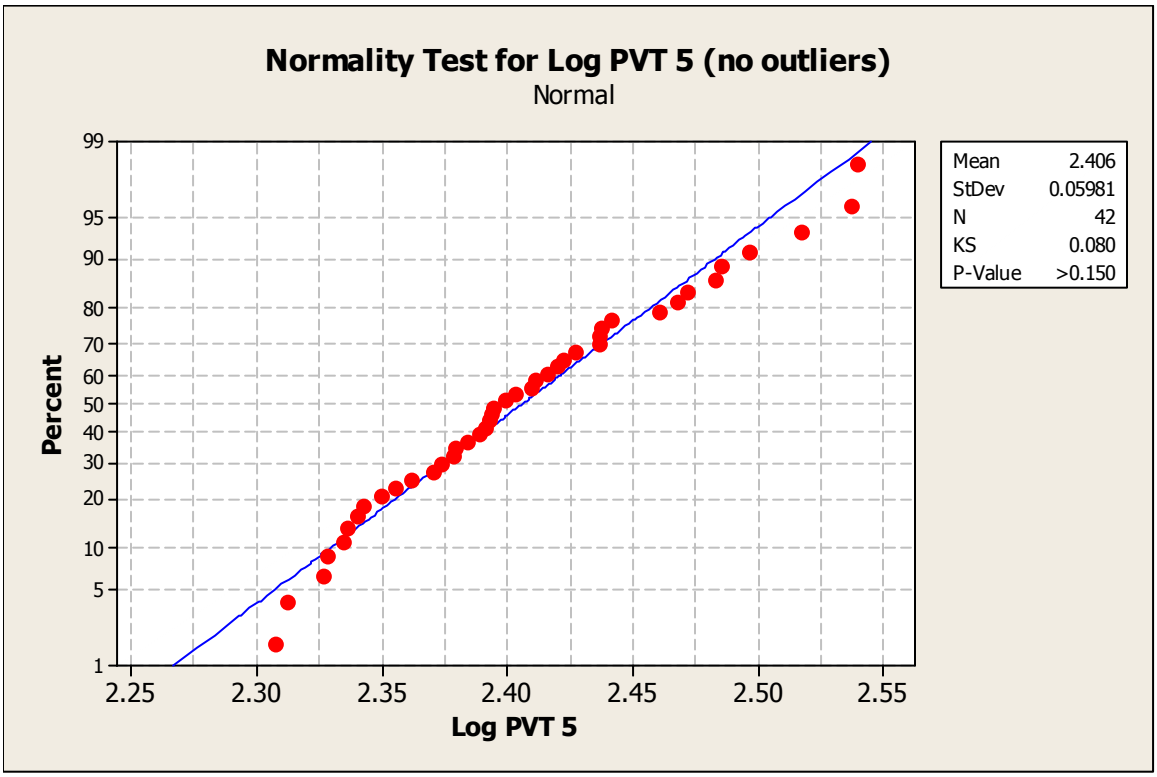


Figure D.15.4 Normality Test for Baseline PVT Log Transform (no outliers)

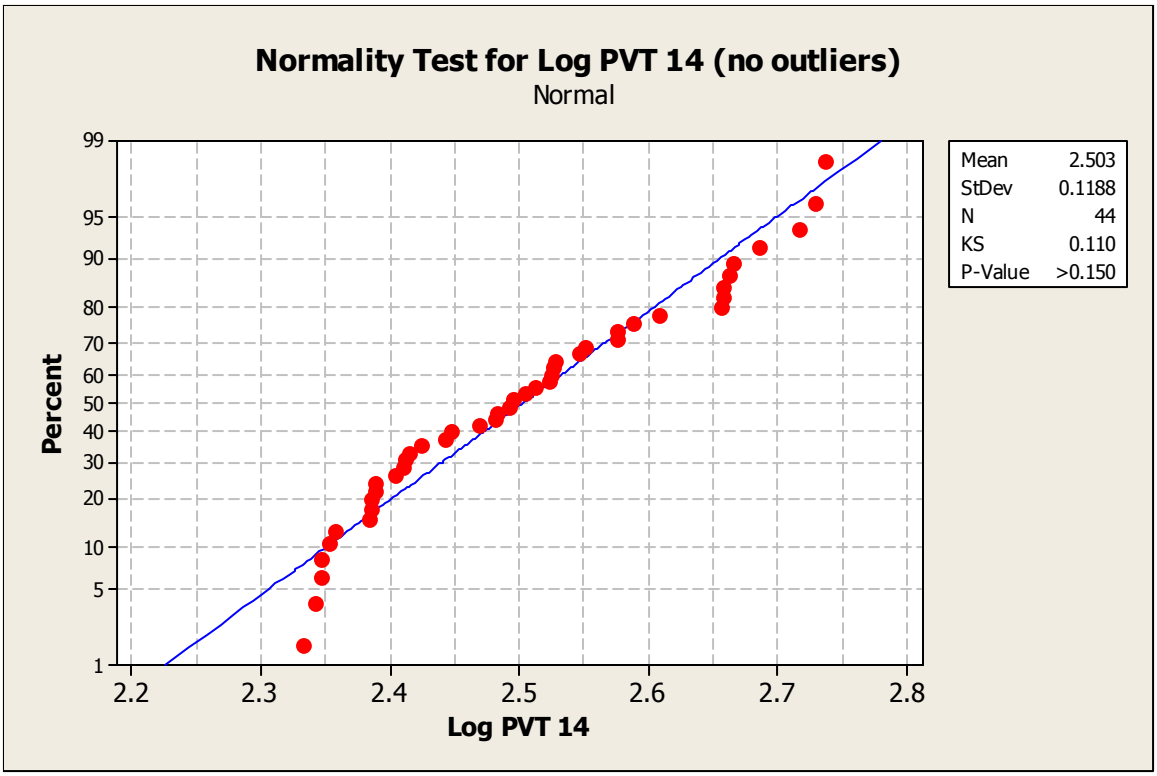


Figure D.15.5 Normality Test for Sleep Deprived PVT Log Transform (no outliers)

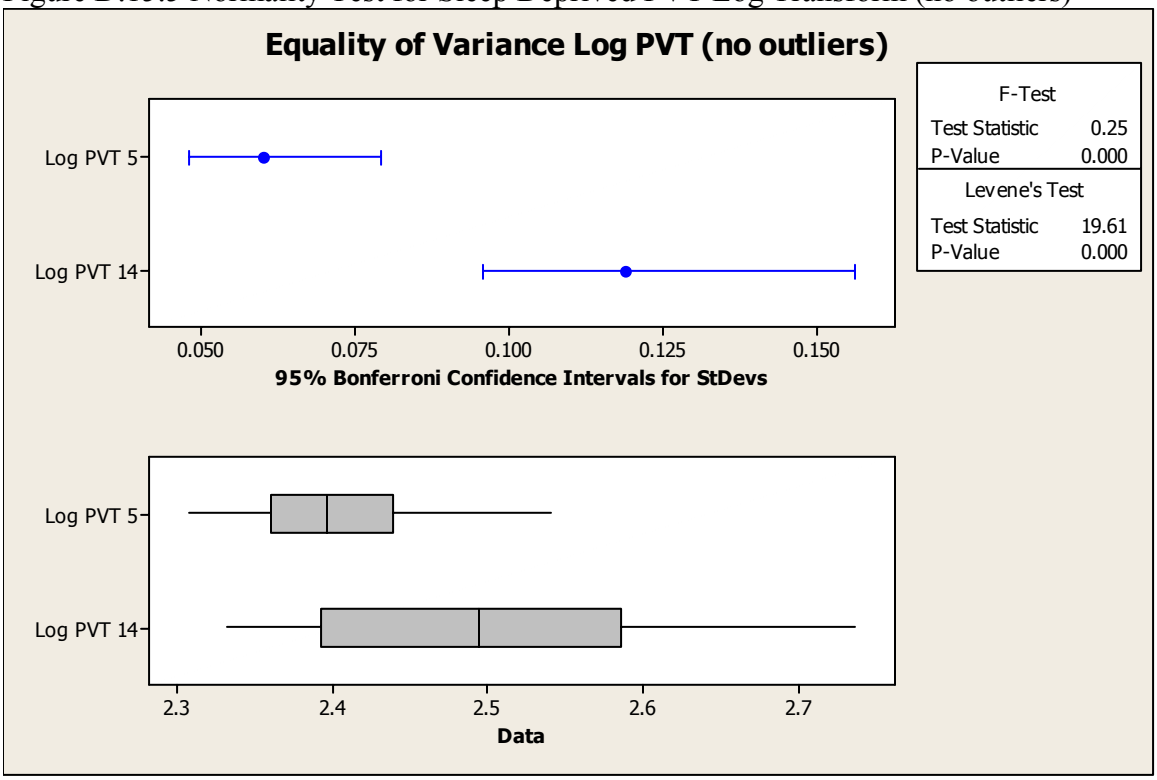


Figure D.15.6 Equality of Variance for PVT Log Transform (no outliers)

D.16 MTT Error Data (Normality and Equality of Variance)

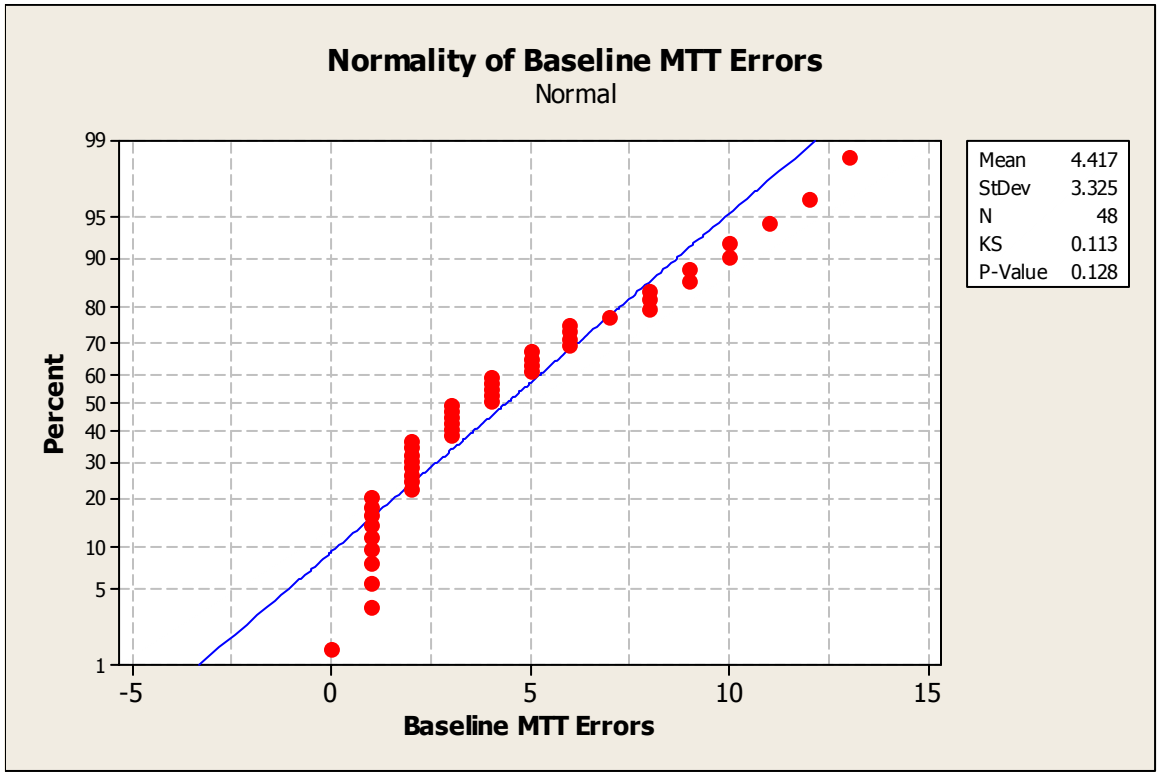


Figure D.16.1 Normality Test for Baseline MTT Errors

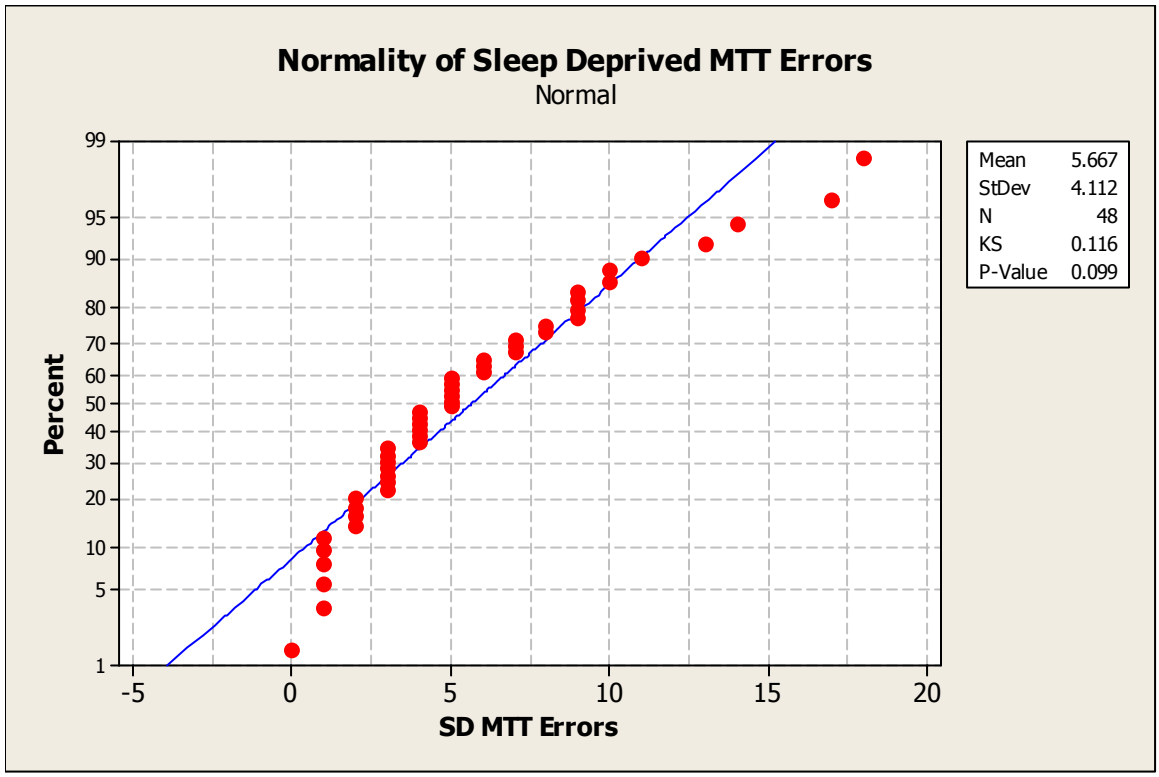


Figure D.16.2 Normality Test for Sleep Deprived MTT Errors

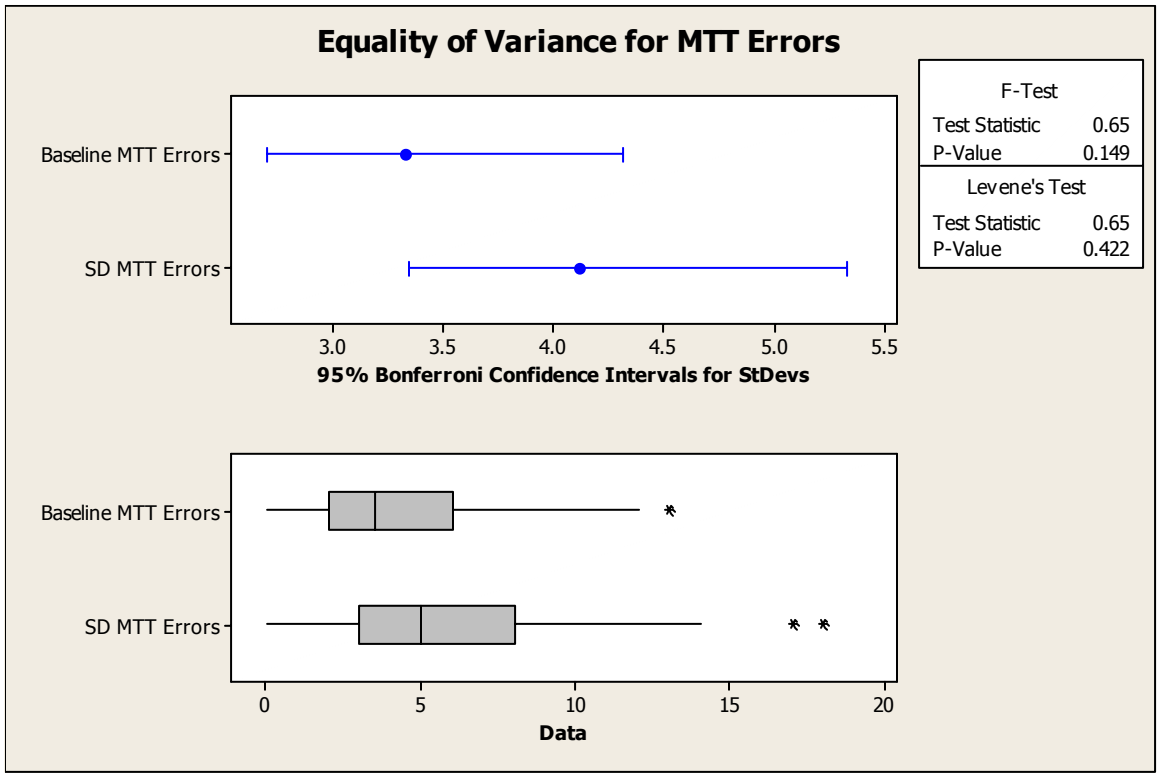


Figure D.16.3 Equality of Variance for MTT Errors

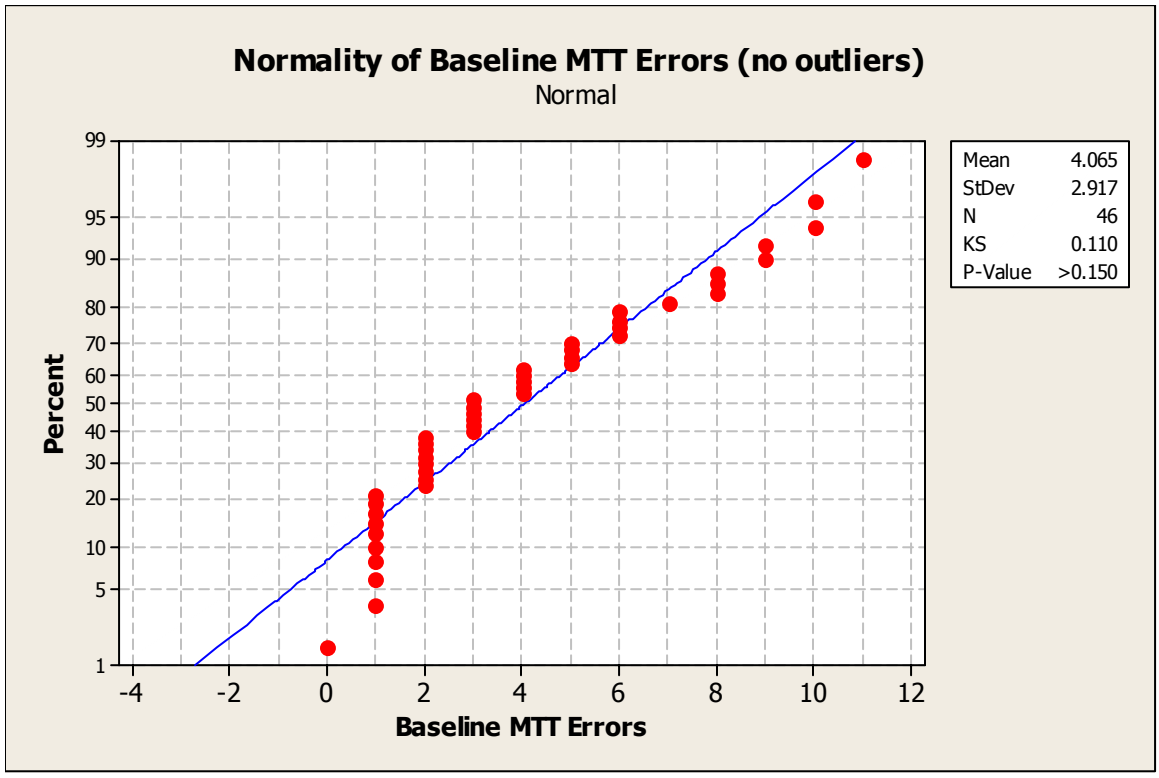


Figure D.16.4 Normality Test for Baseline MTT Errors (no outliers)

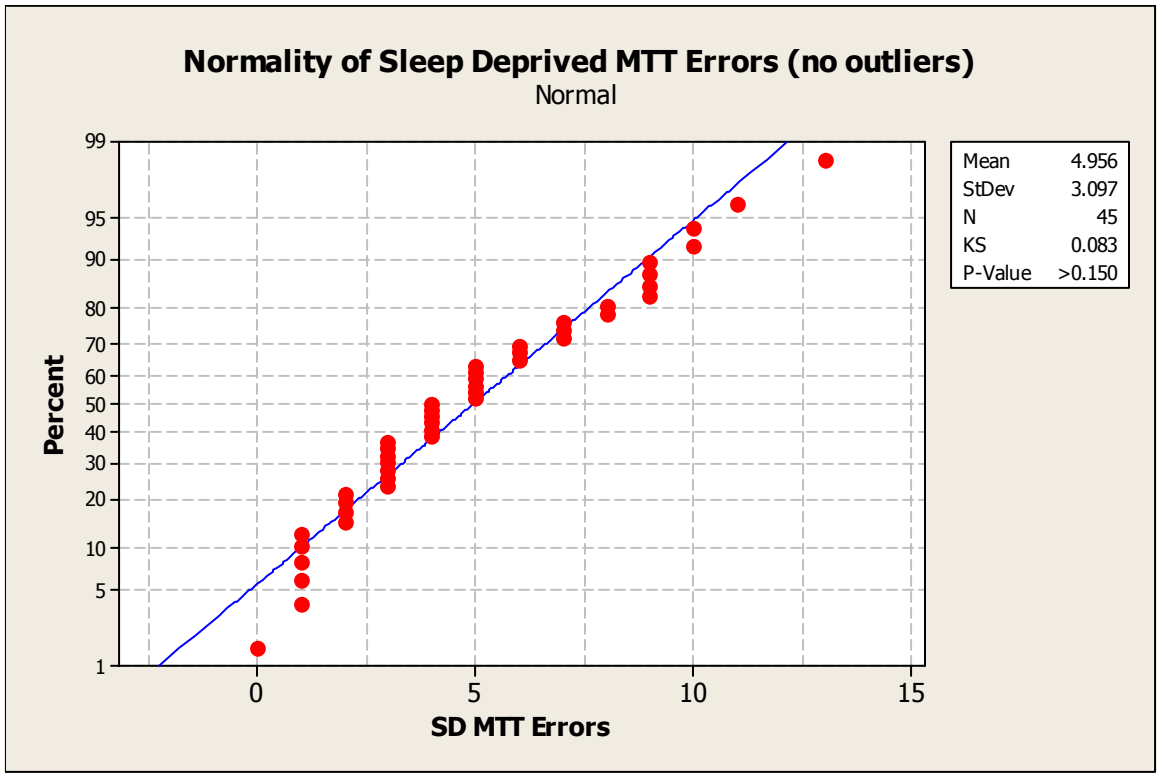


Figure D.16.5 Normality Test for Sleep Deprived MTT Errors (no outliers)

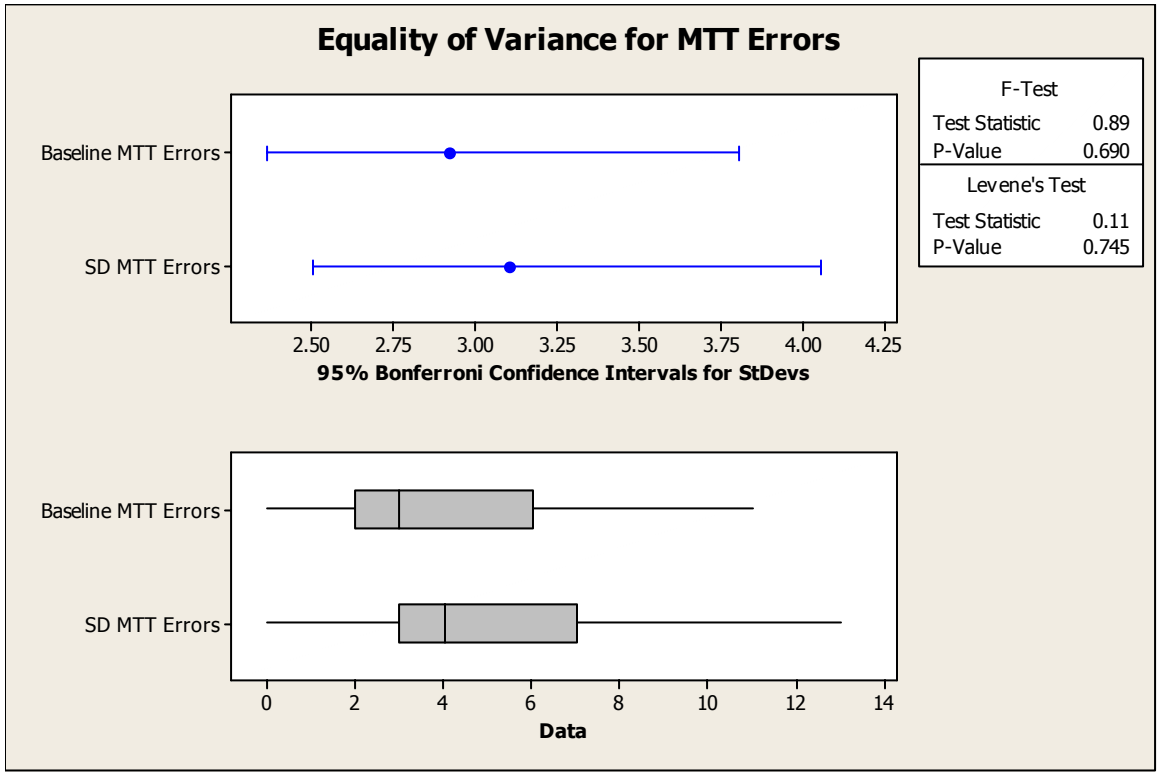


Figure D.16.6 Equality of Variance for MTT Errors (no outliers)

D.17 LS Net Score Data (Normality and Equality of Variance)

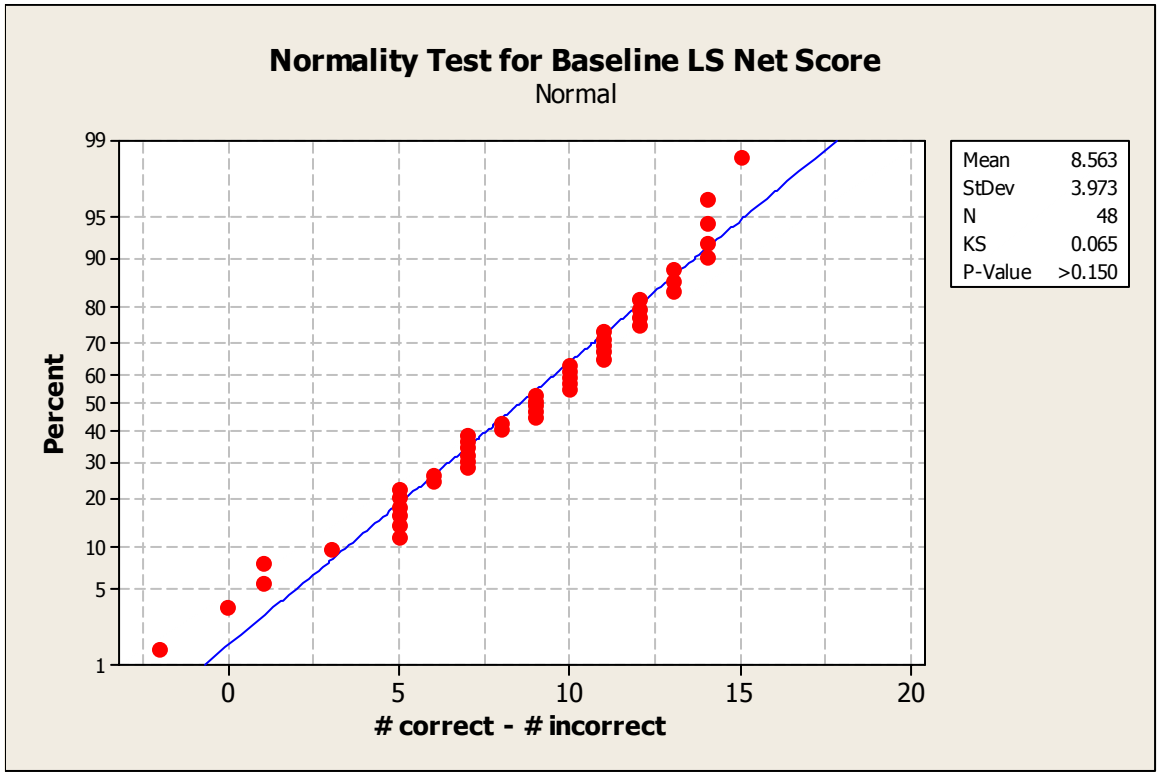


Figure D.17.1 Normality Test for Baseline LS Net Score

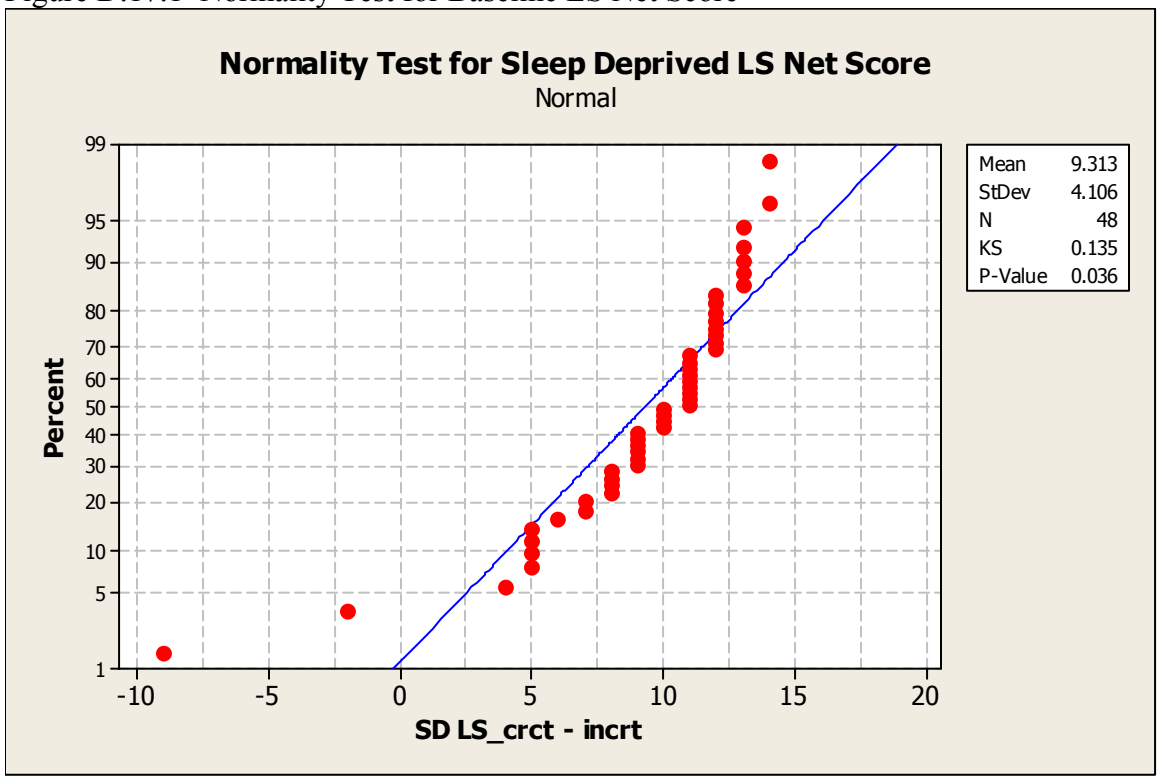


Figure D.17.2 Normality Test for Sleep Deprived LS Net Score

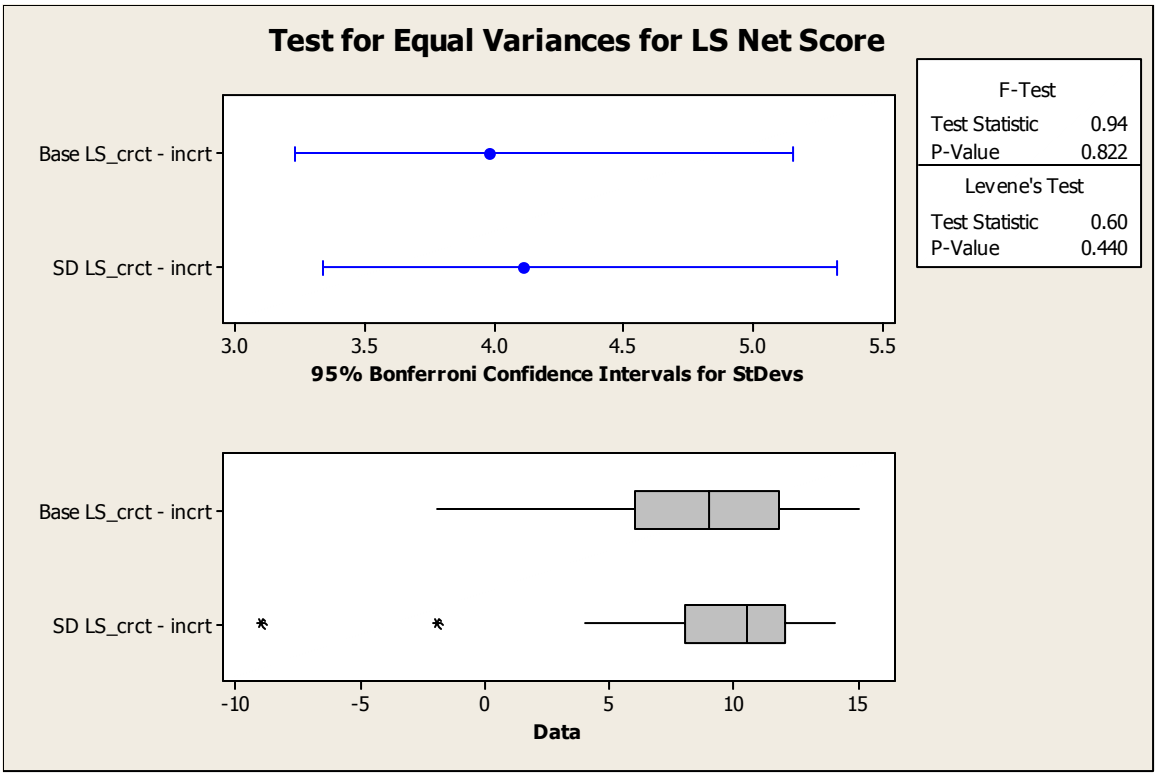


Figure D.17.3 Equality of Variance for LS Net Score

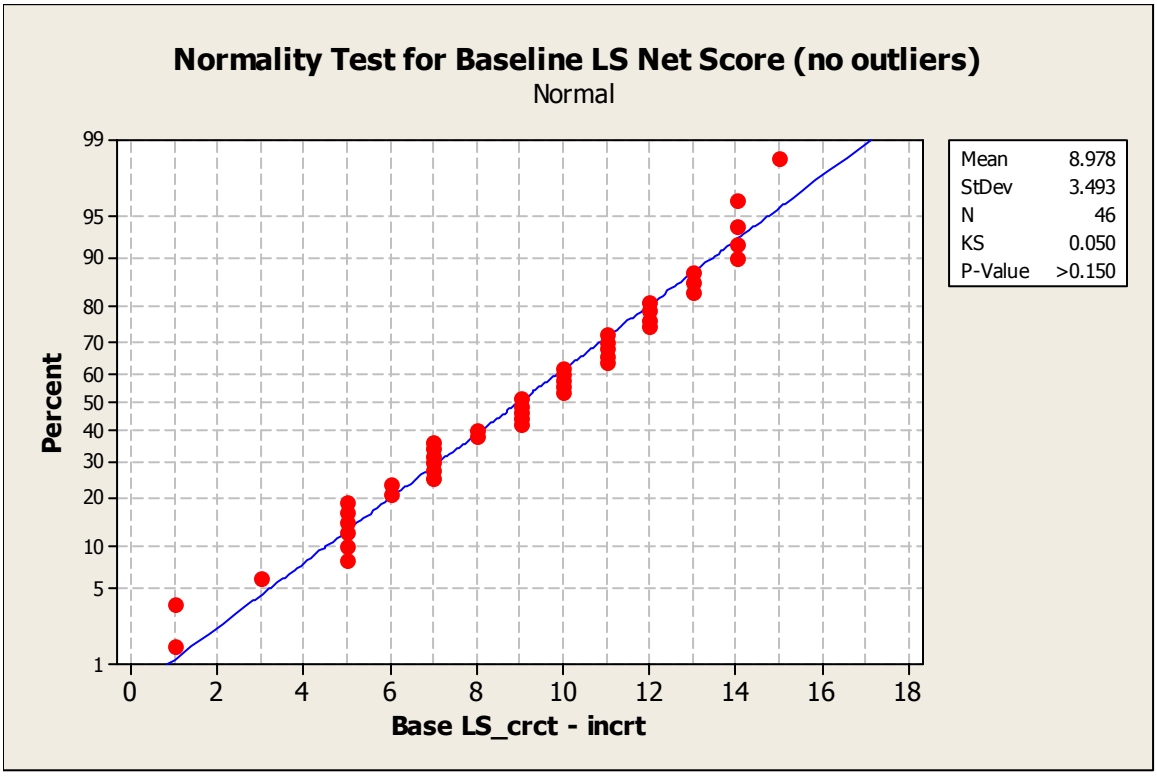


Figure D.17.4 Normality Test for Baseline LS Net Score (no outliers)

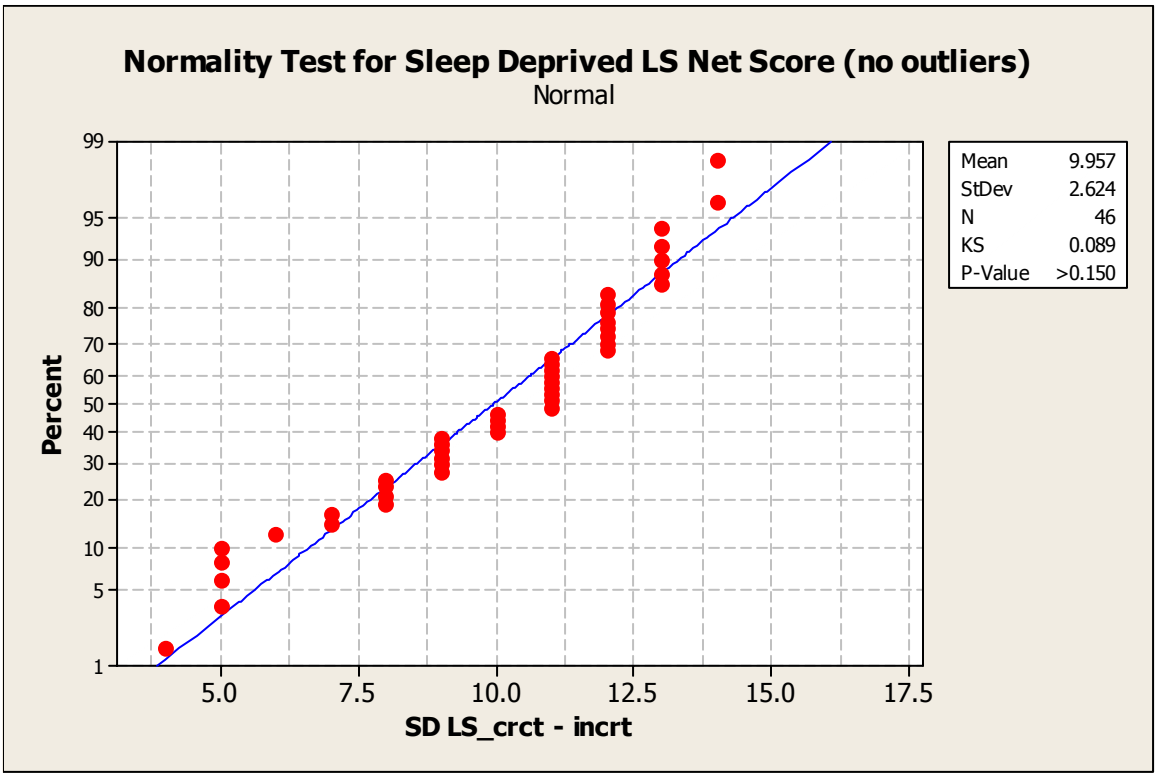


Figure D.17.5 Normality Test for Sleep Deprived LS Net Score (no outliers)

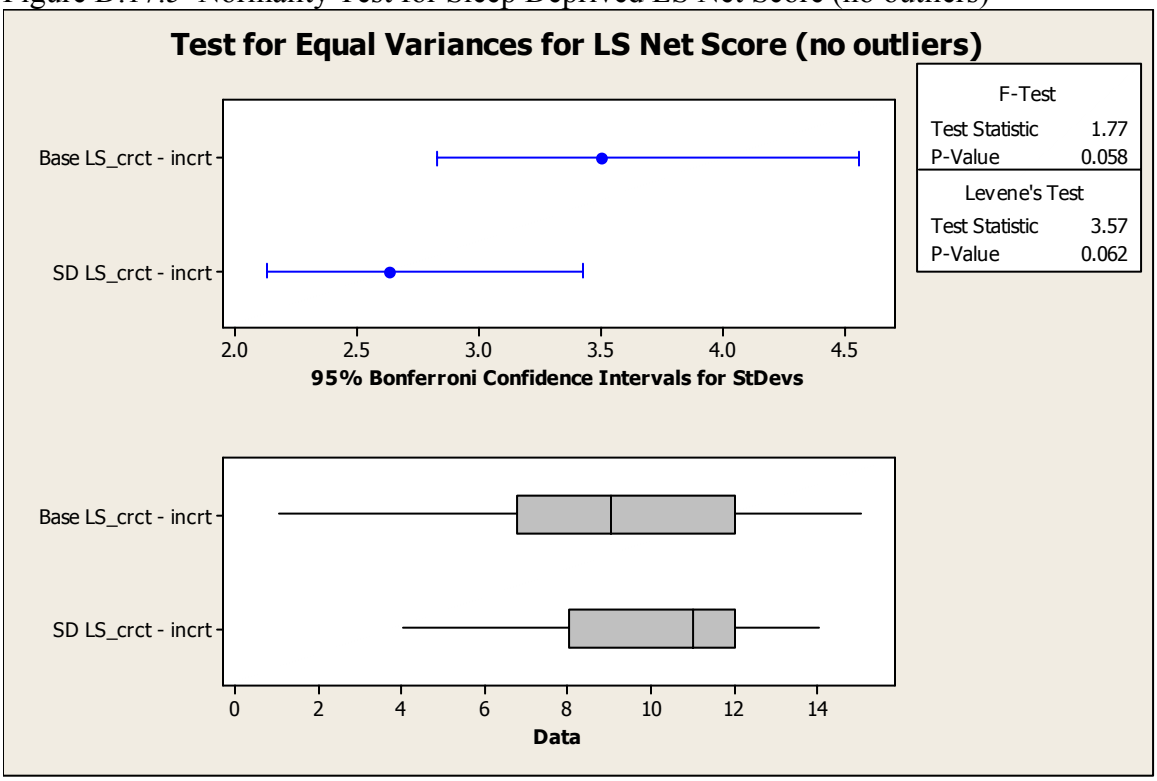


Figure D.17.6 Equality of Variance for LS Net Score (no outliers)

D.18 NS Net Score Data (Normality and Equality of Variance)

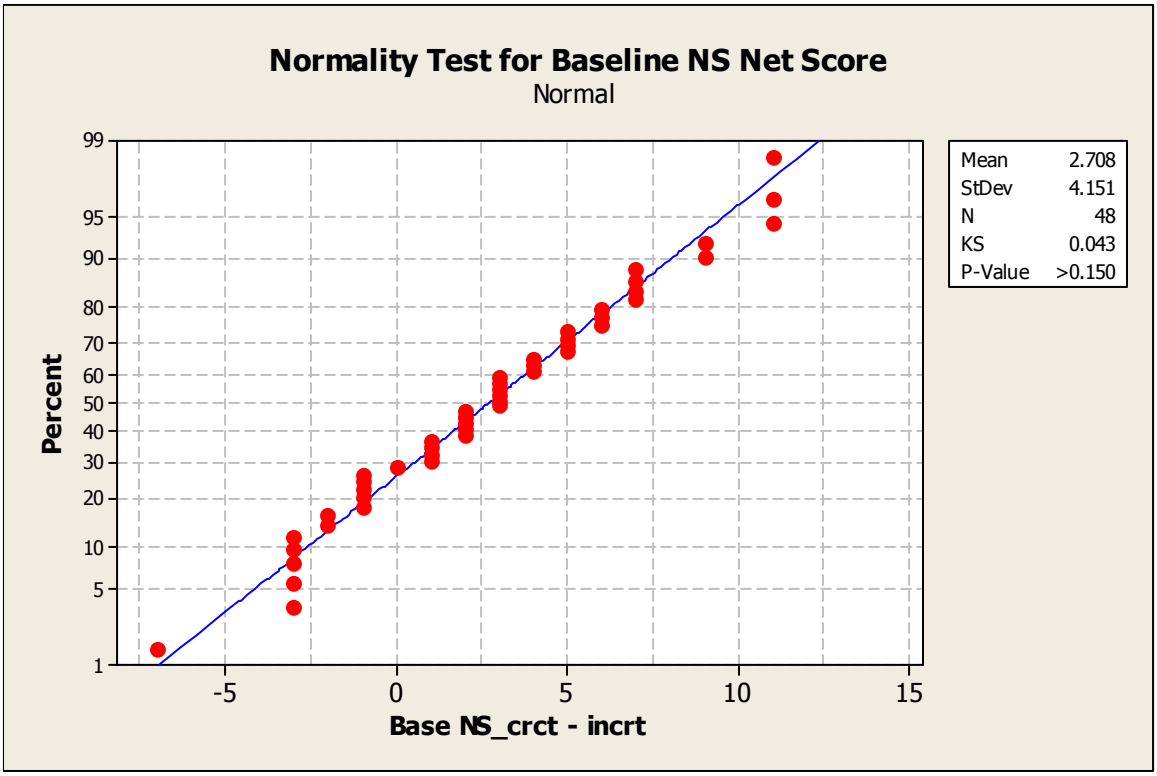


Figure D.18.1 Normality Test for Baseline NS Net Score

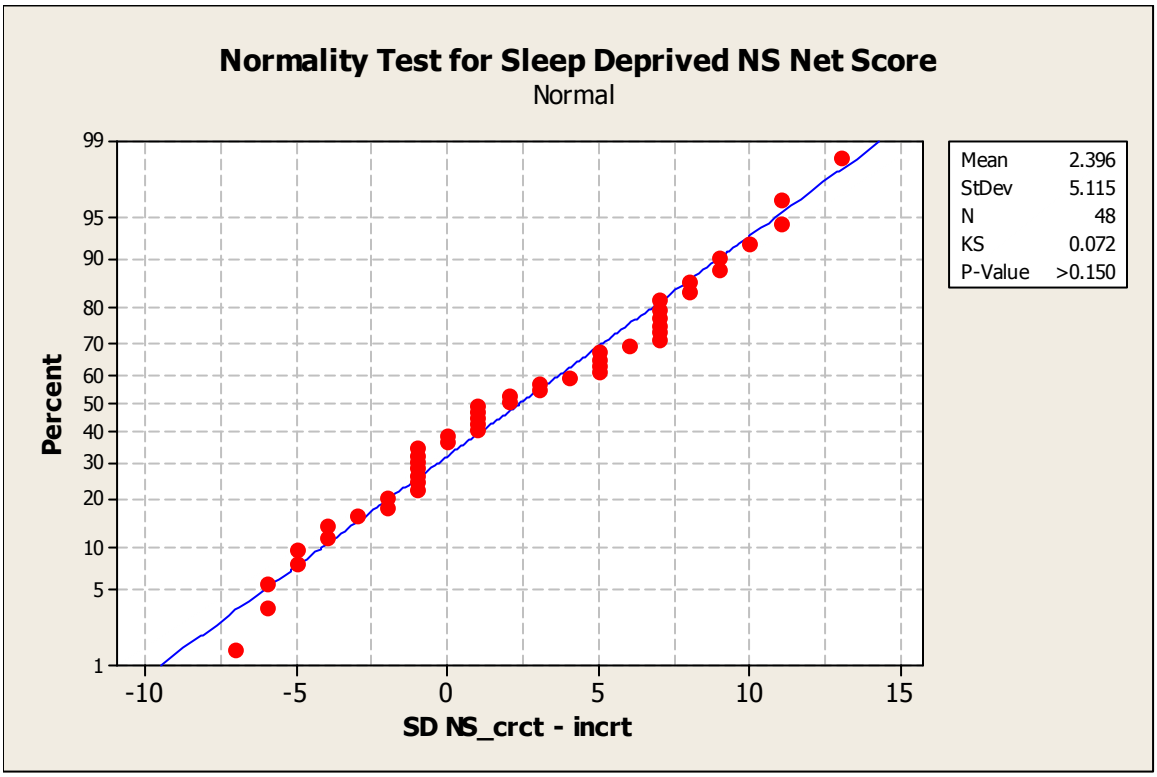


Figure D.18.2 Normality Test for Sleep Deprived NS Net Score

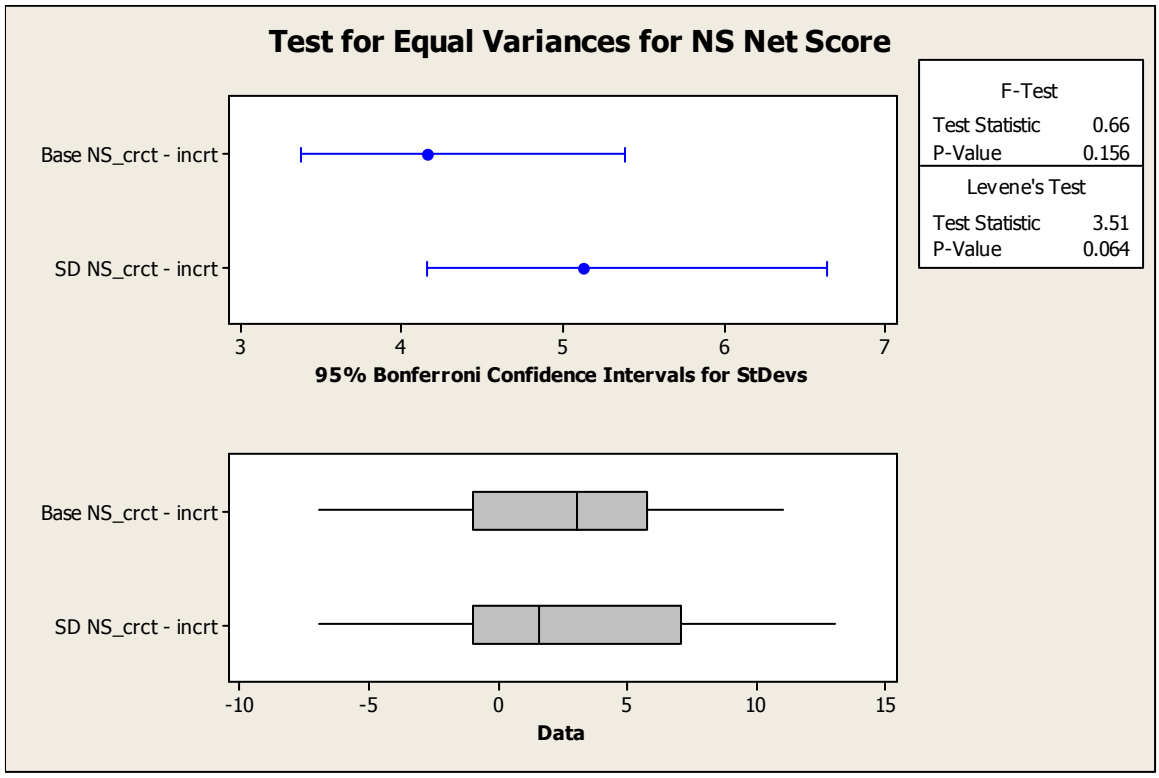


Figure D.18.3 Equality of Variance for NS Net Score

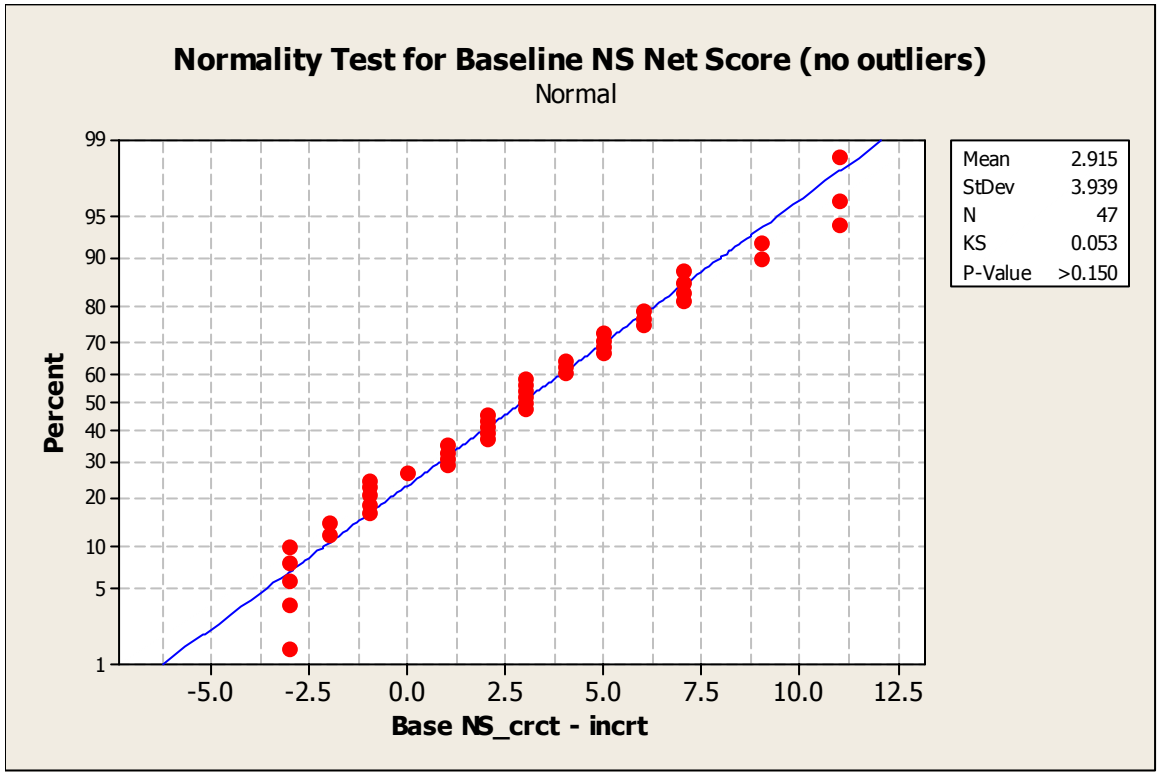


Figure D.18.4 Normality Test for Baseline NS Net Score (no outliers)

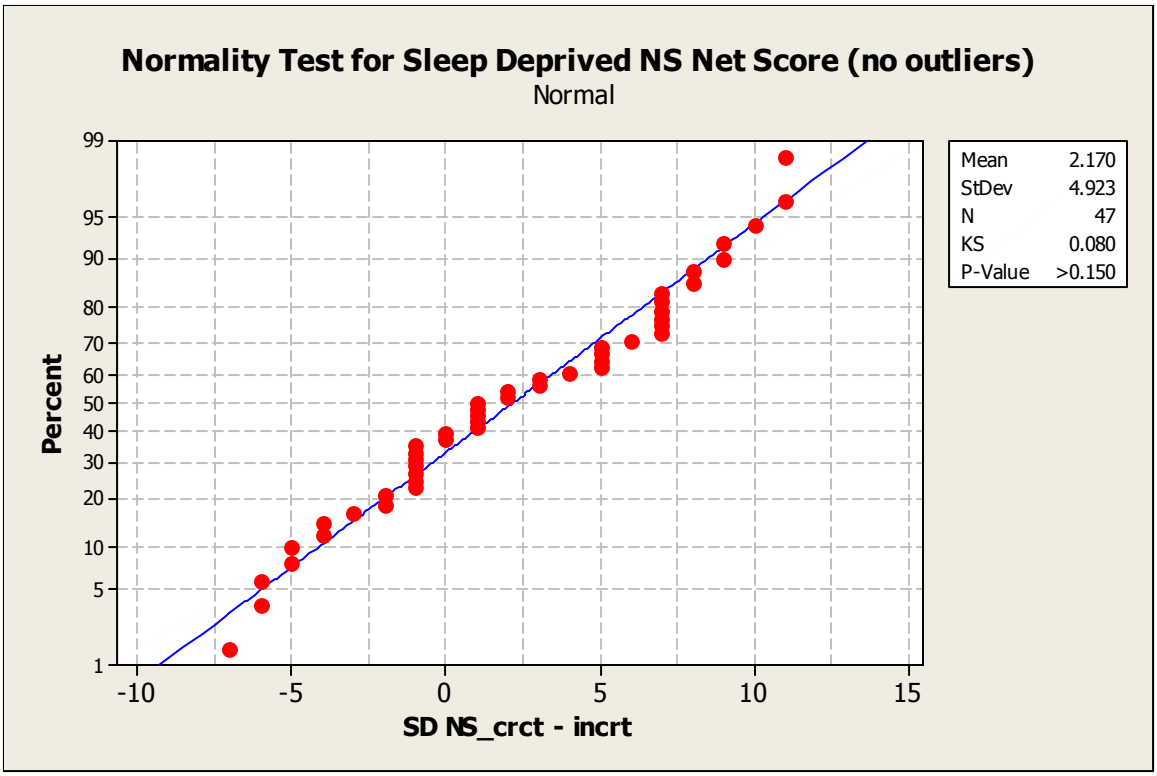


Figure D.18.5 Normality Test for Sleep Deprived NS Net Score (no outliers)

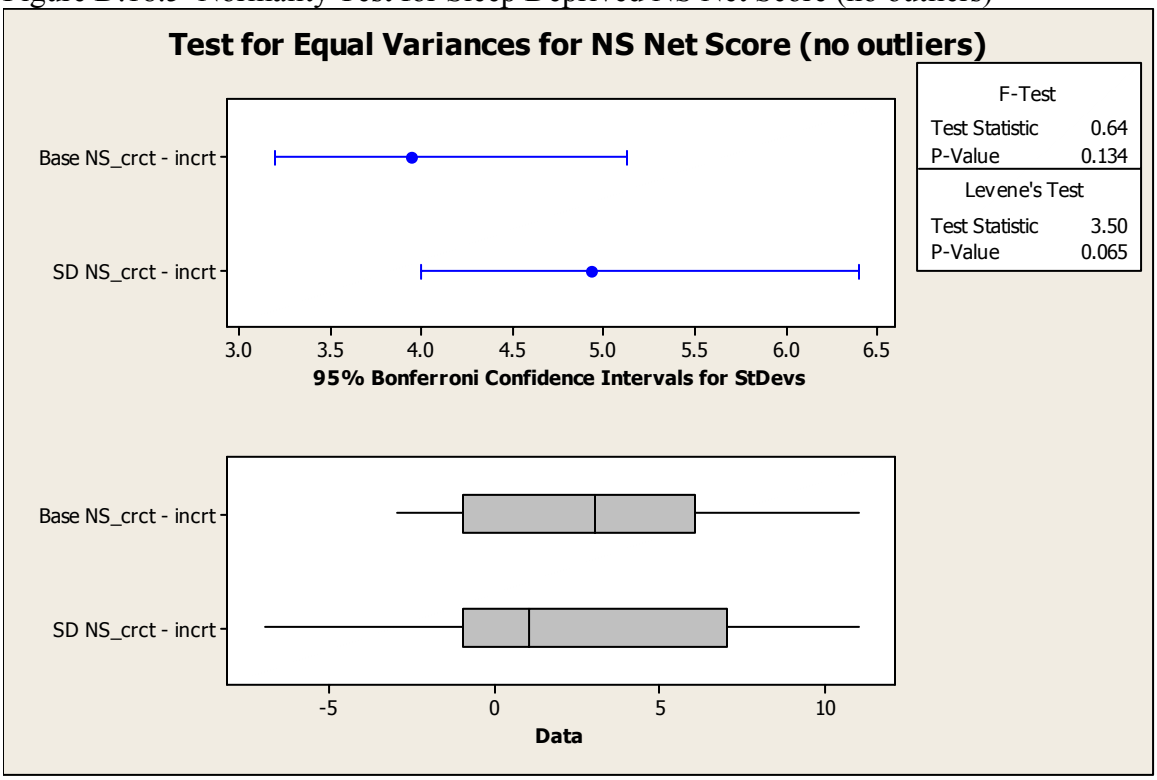


Figure D.18.6 Equality of Variance Test for NS Net Score (no outliers)

APPENDIX E. PAIRED T-TESTS

E.1 Paired T-Test Assumptions (Normality of Paired Differences)

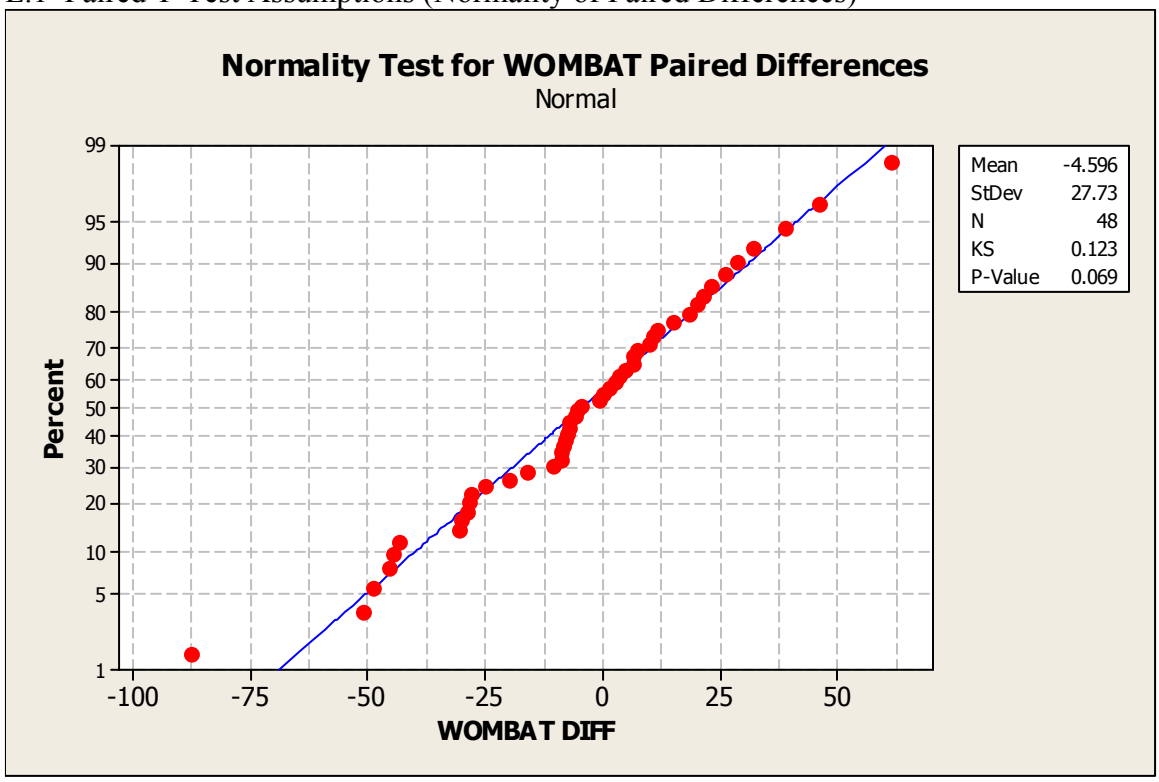


Figure E.1 Normality test for WOMBAT paired differences

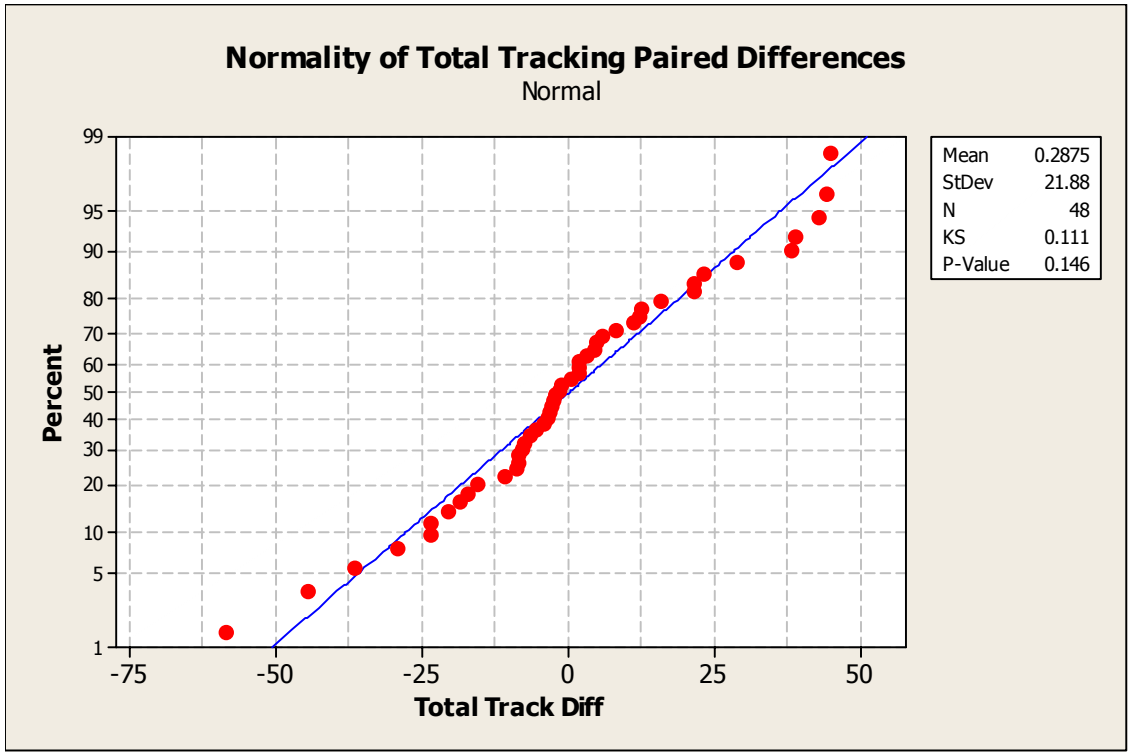


Figure E.2 Normality test for Total Tracking paired differences

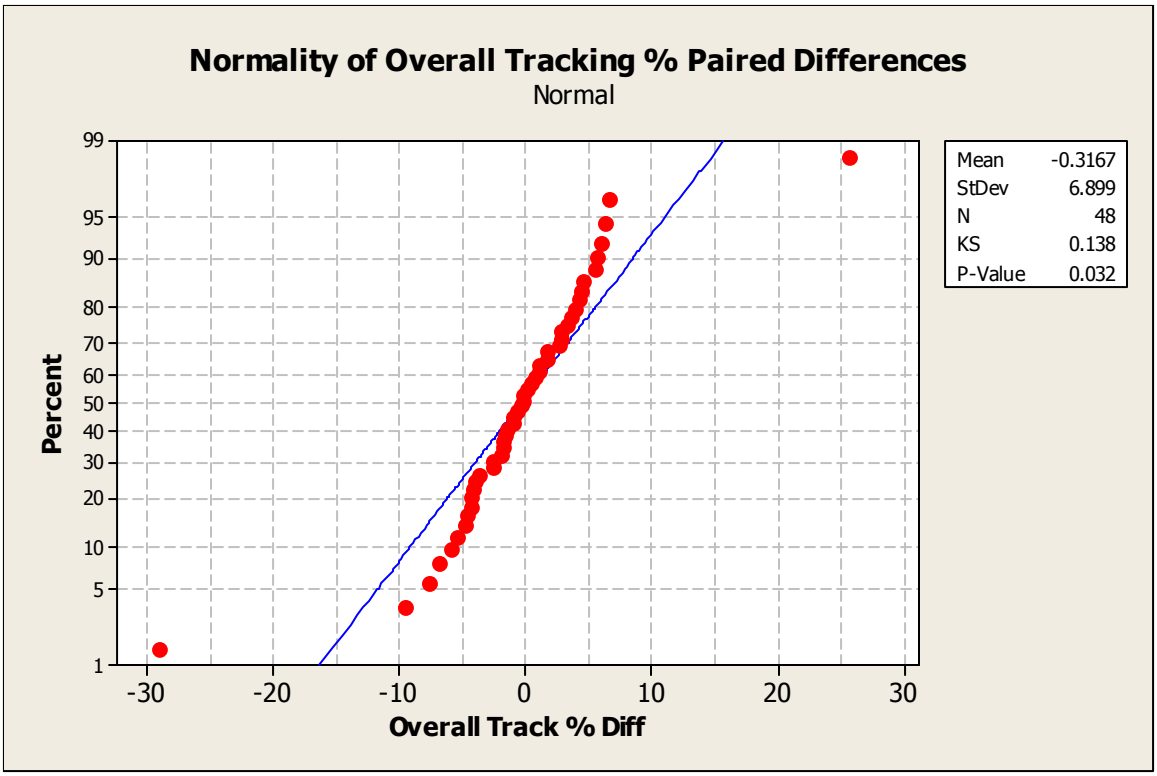


Figure E.3 Normality test for Overall Tracking Percent paired differences

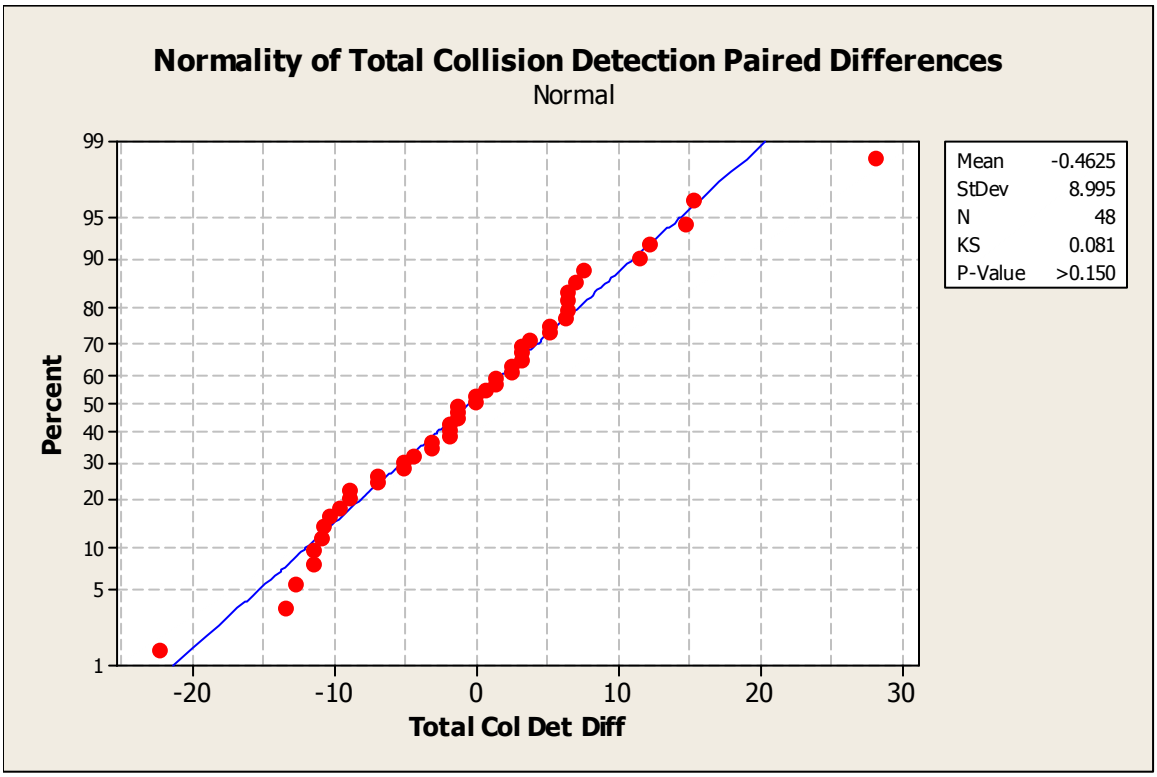


Figure E.4 Normality test for Total Collision Detection paired differences

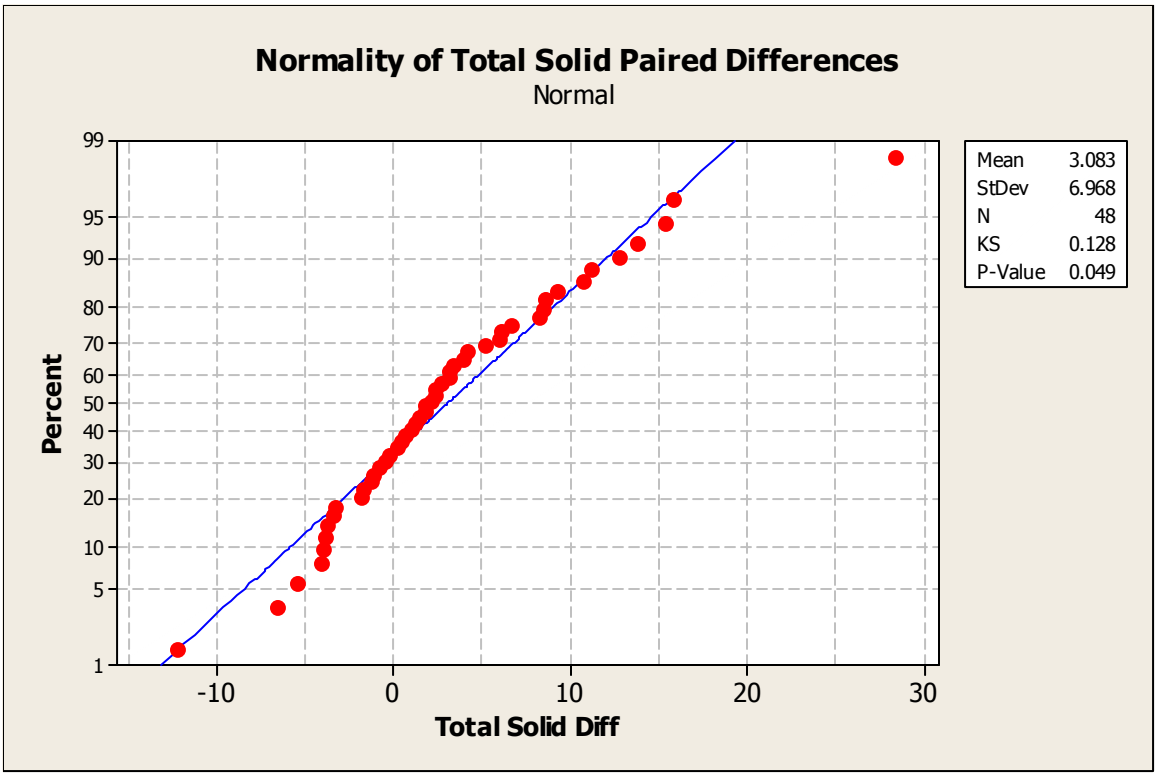


Figure E.5 Normality test for Total Solid paired differences

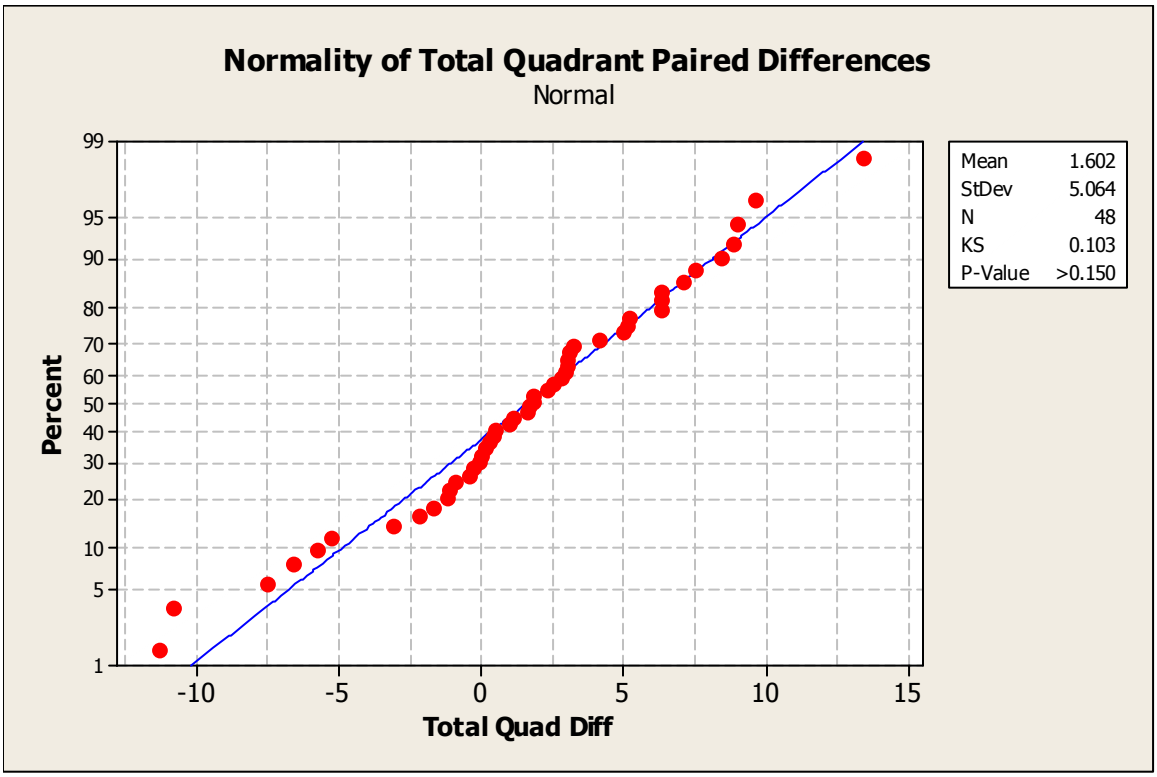


Figure E.6 Normality test for Total Quadrant paired differences

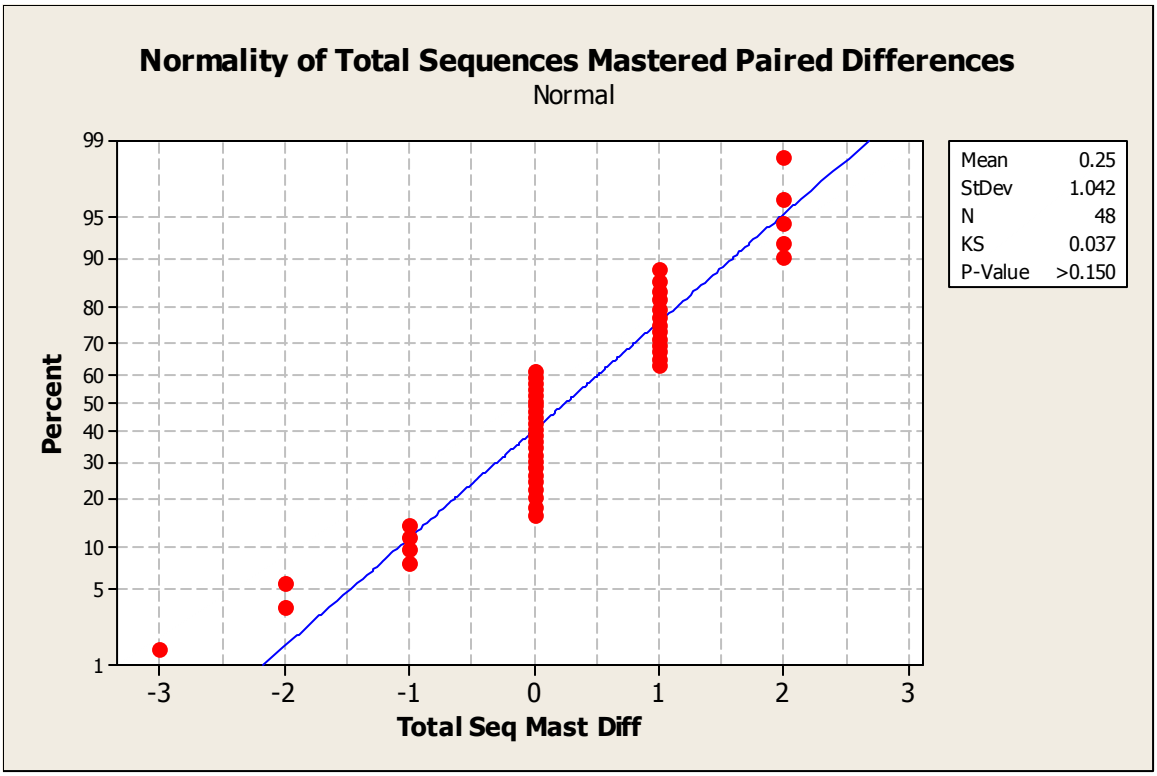


Figure E.7 Normality test for Total Sequences Mastered paired differences

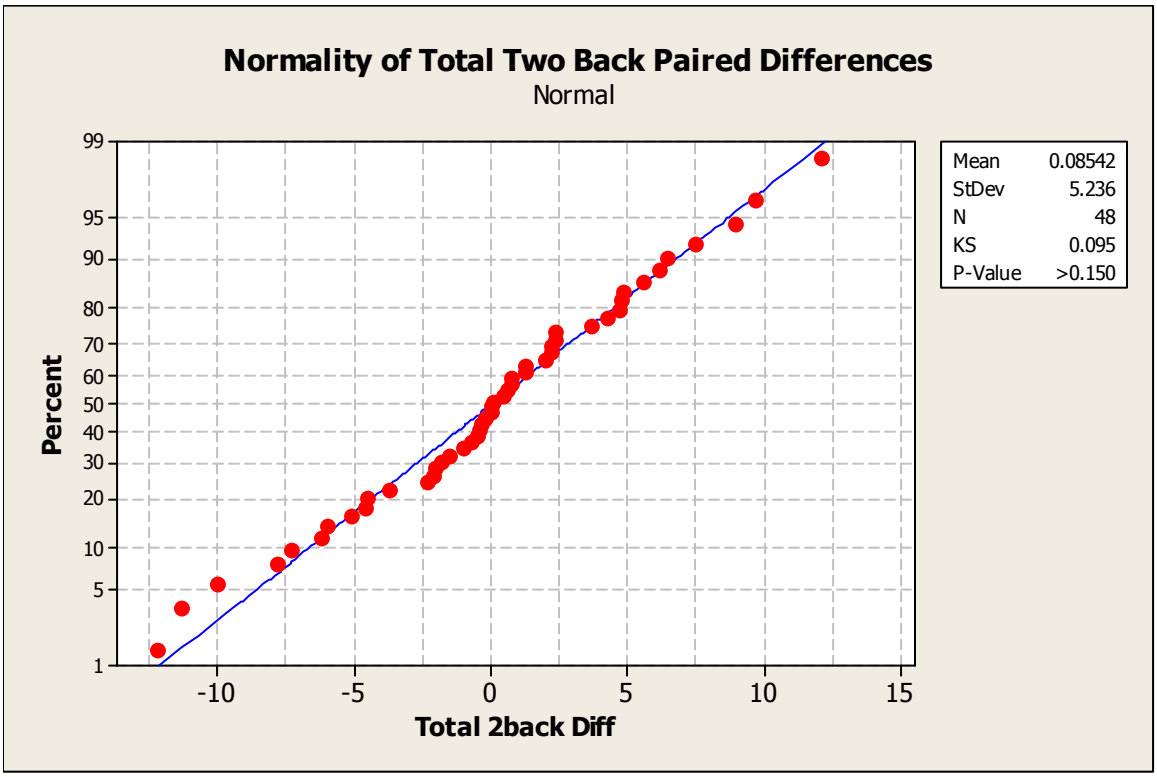


Figure E.8 Normality test for Total Two-Back paired differences

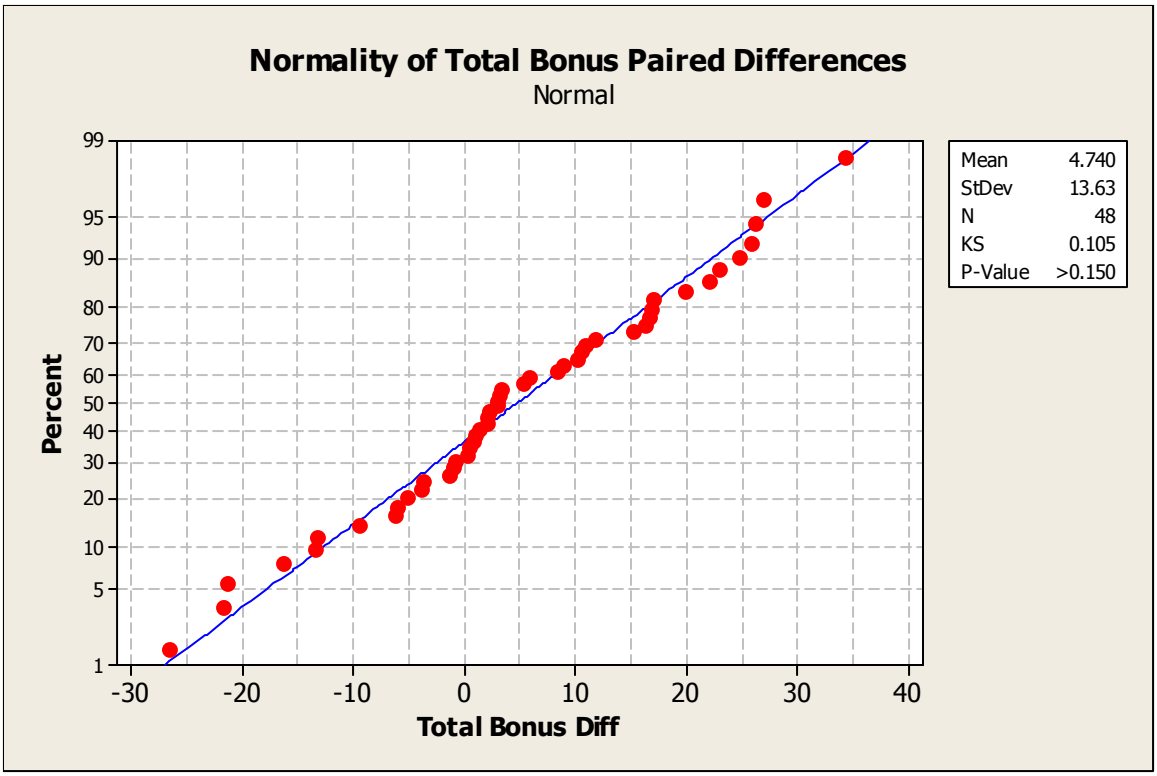


Figure E.9 Normality test for Total Bonus paired differences

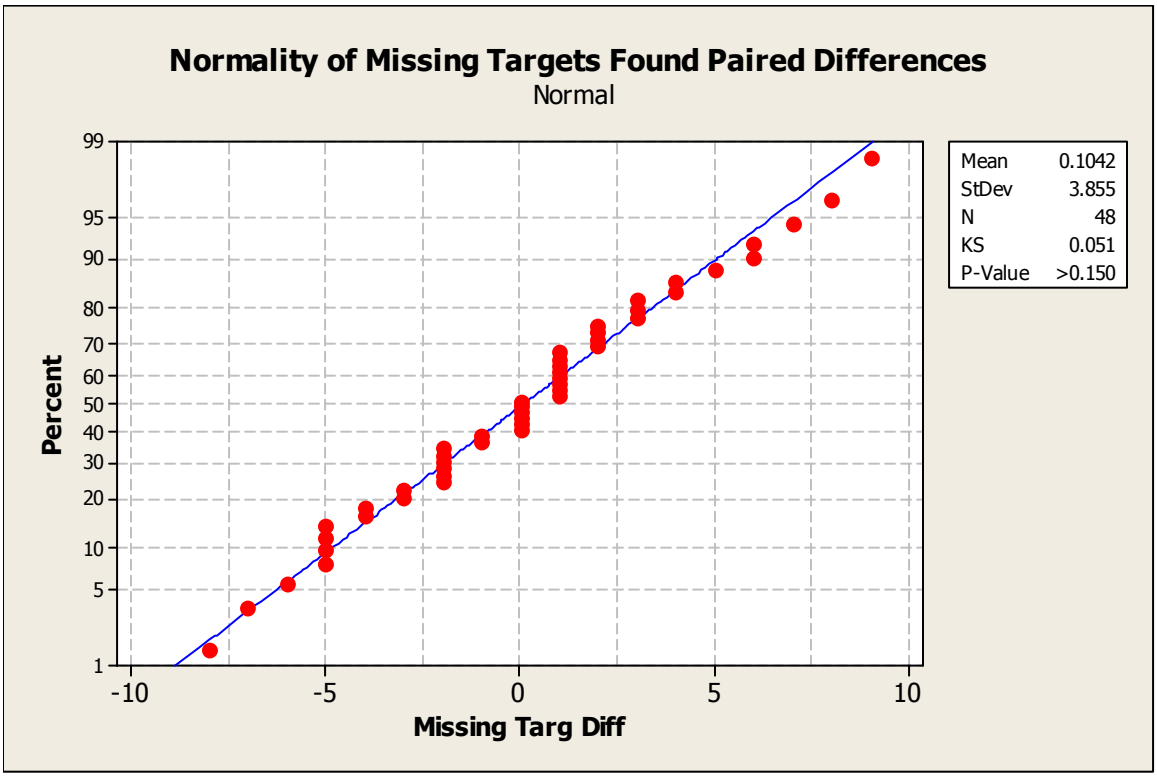


Figure E.10 Normality test for Missing Targets Found paired differences

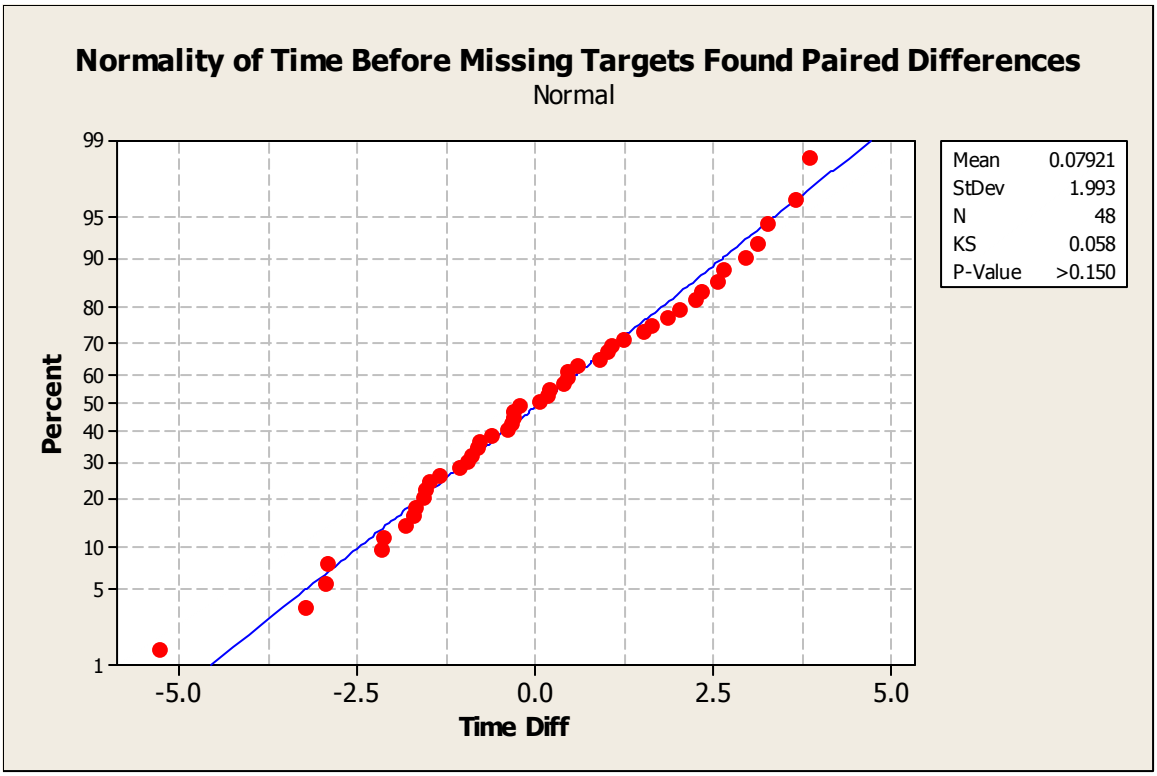


Figure E.11 Normality test for Time Before Missing Target Founc paired differences

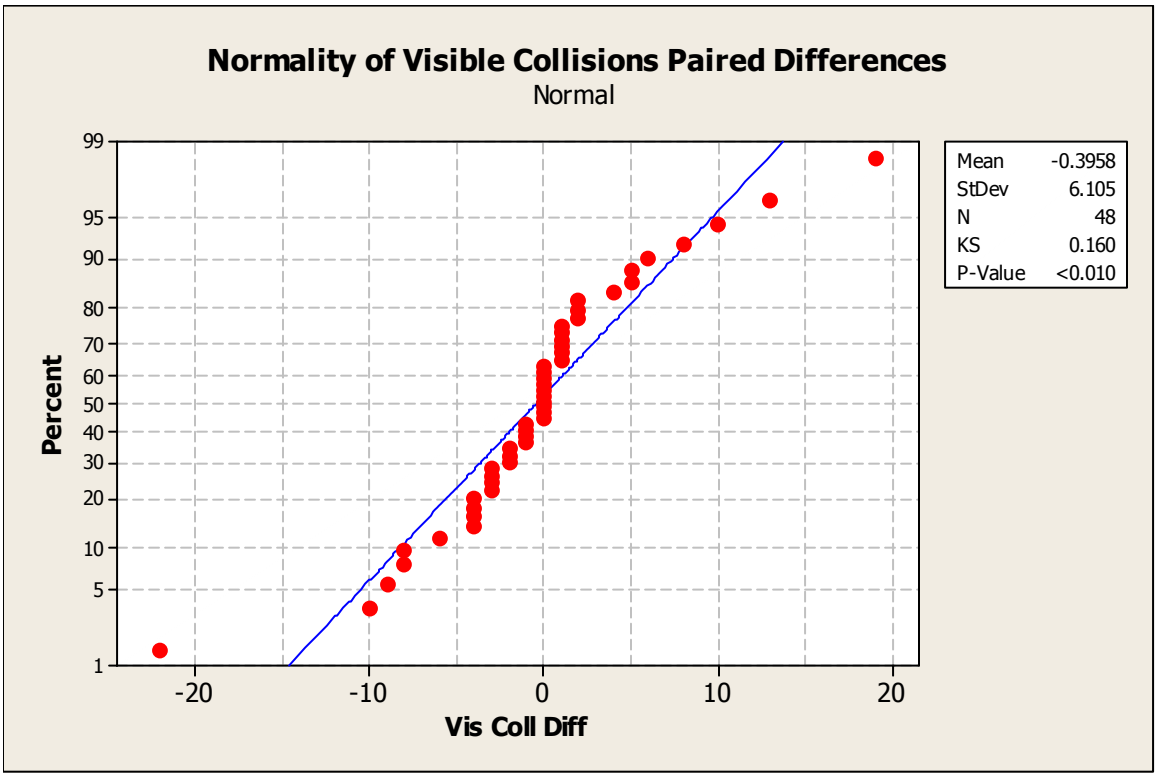


Figure E.12 Normality test for Visible Collisions paired differences

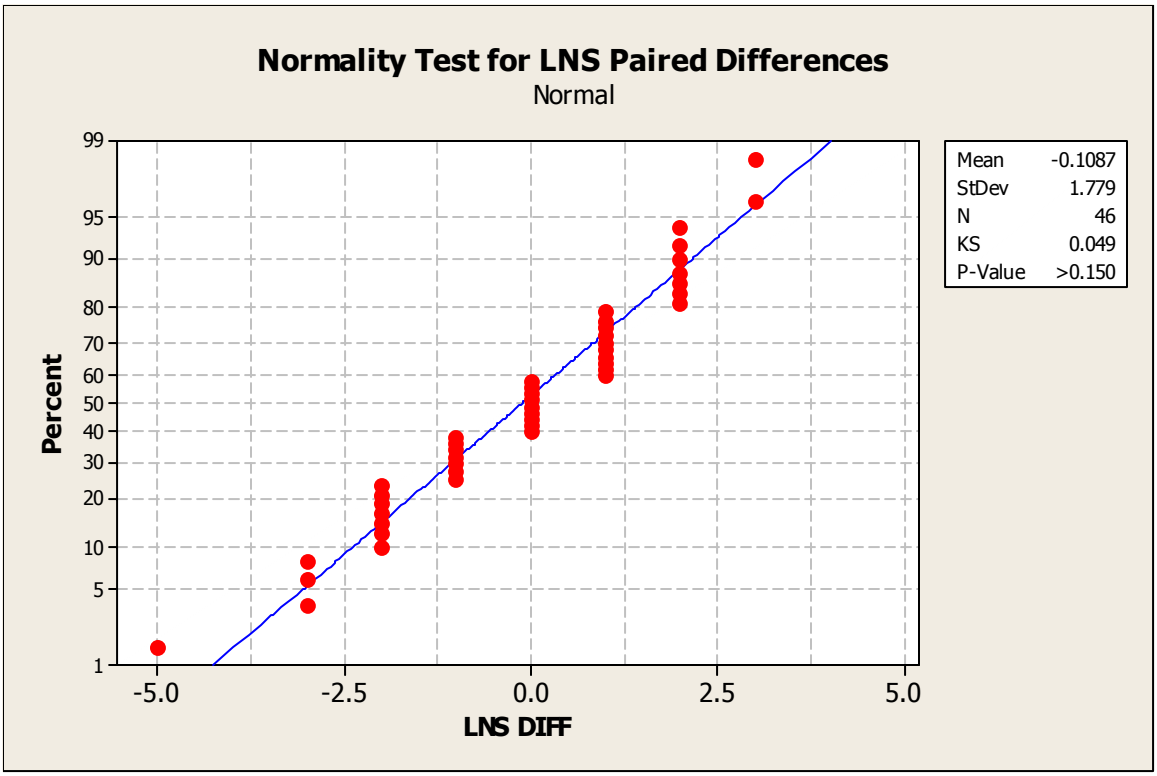


Figure E.13 Normality test for LNS paired differences

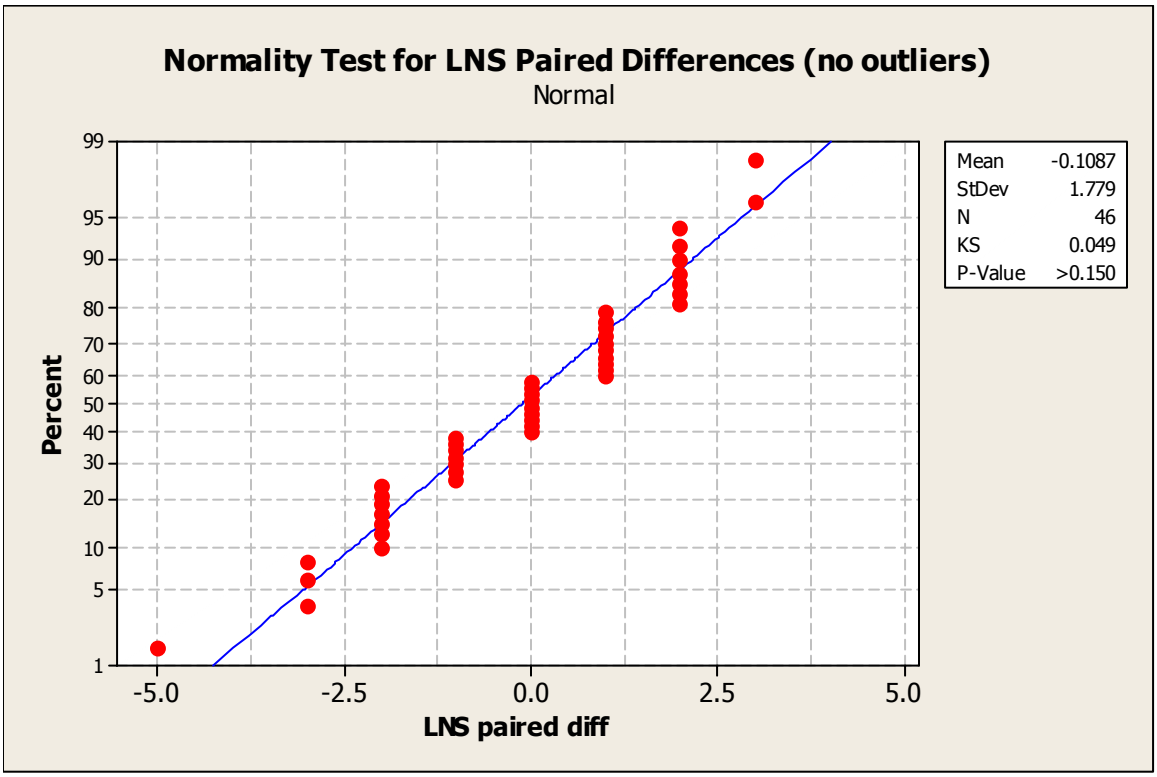


Figure E.14 Normality test for LNS paired differences (no outliers)

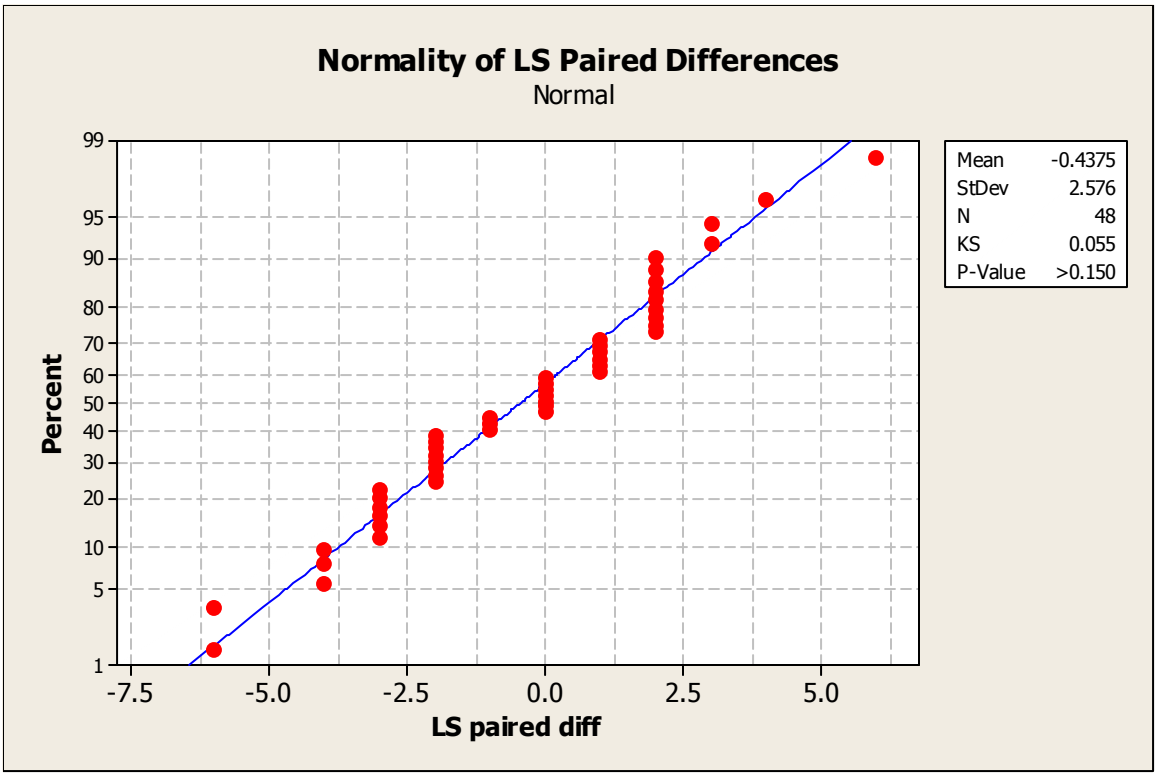


Figure E.15 Normality test for LS paired differences

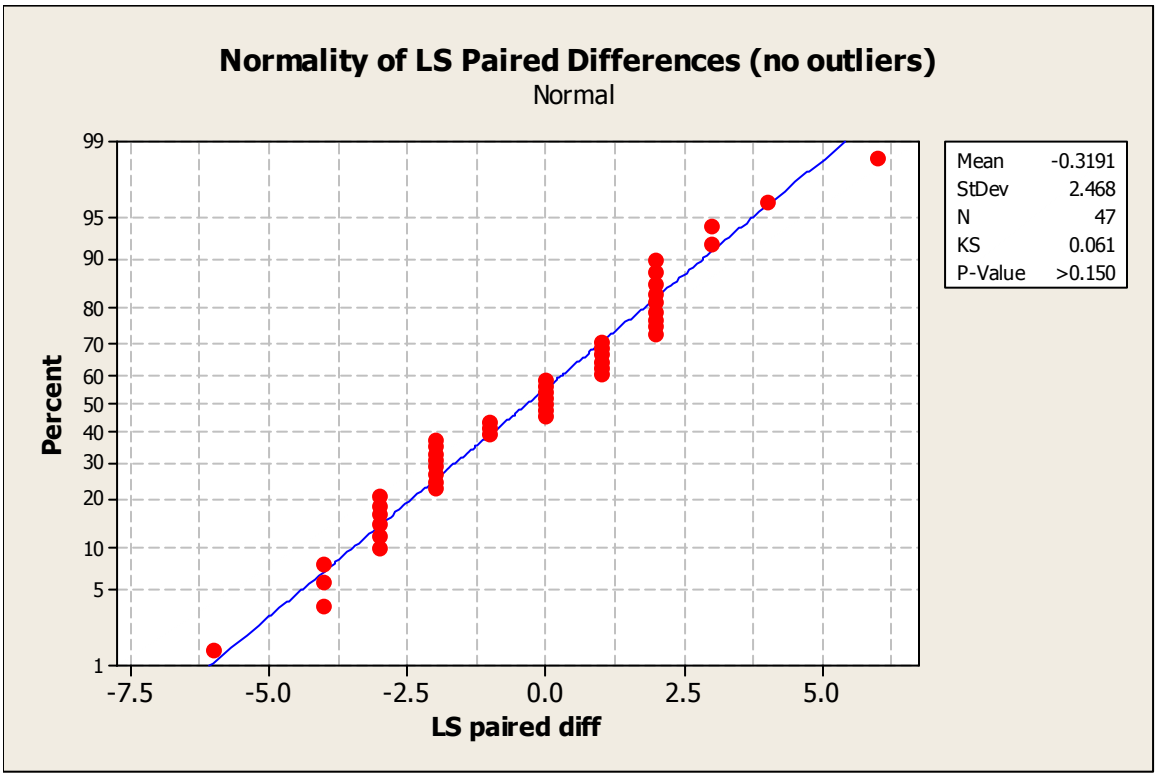


Figure E.16 Normality test for LS paired differences (no outliers)

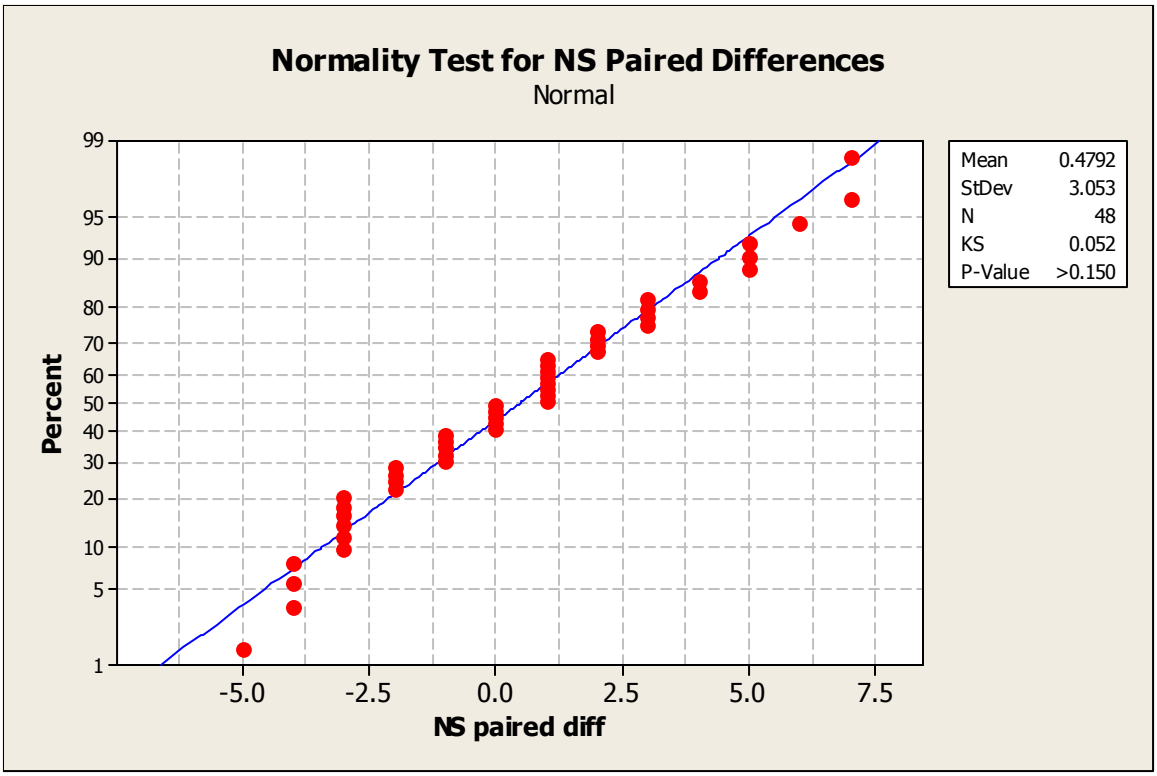


Figure E.17 Normality test for NS paired differences

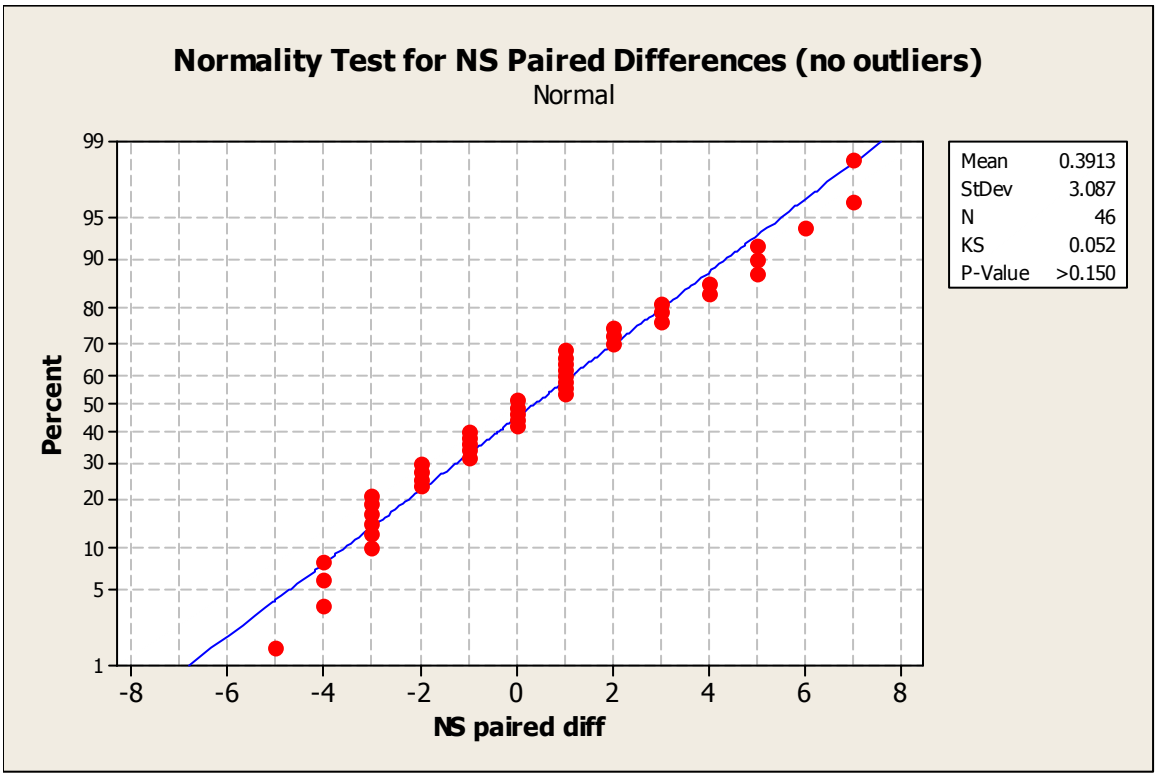


Figure E.18 Normality test for NS paired differences (no outliers)

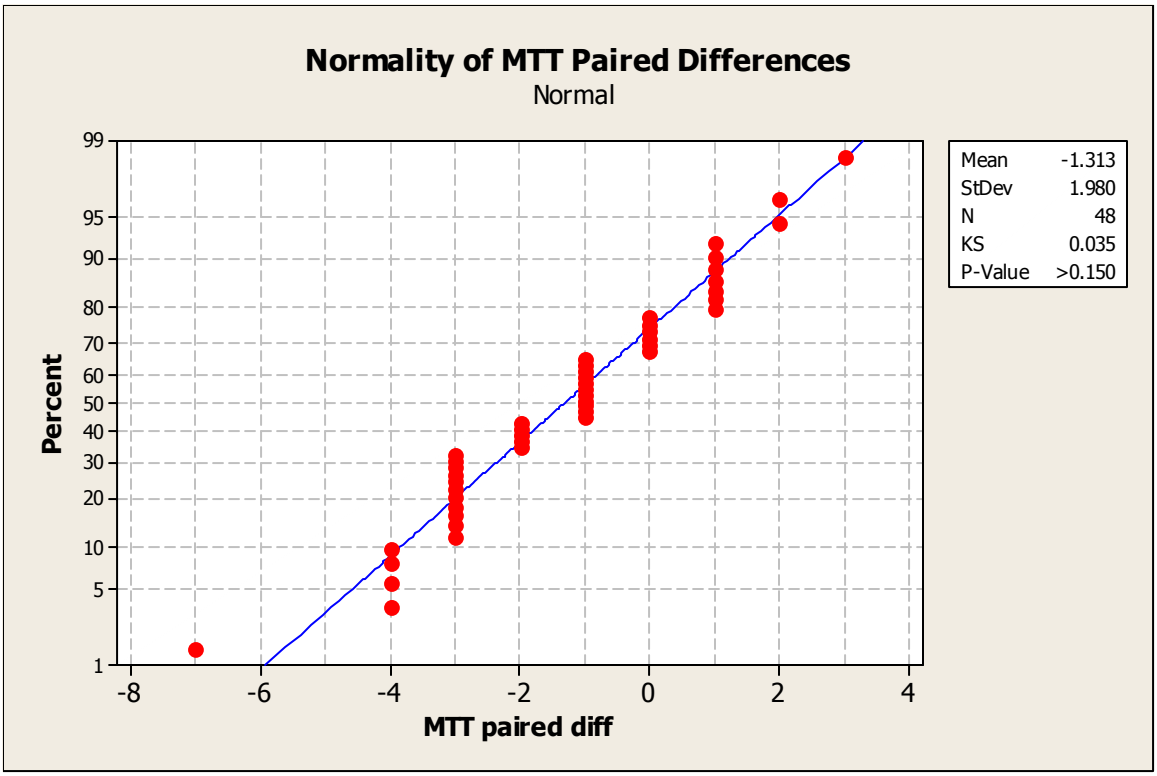


Figure E.19 Normality Test for MTT paired differences

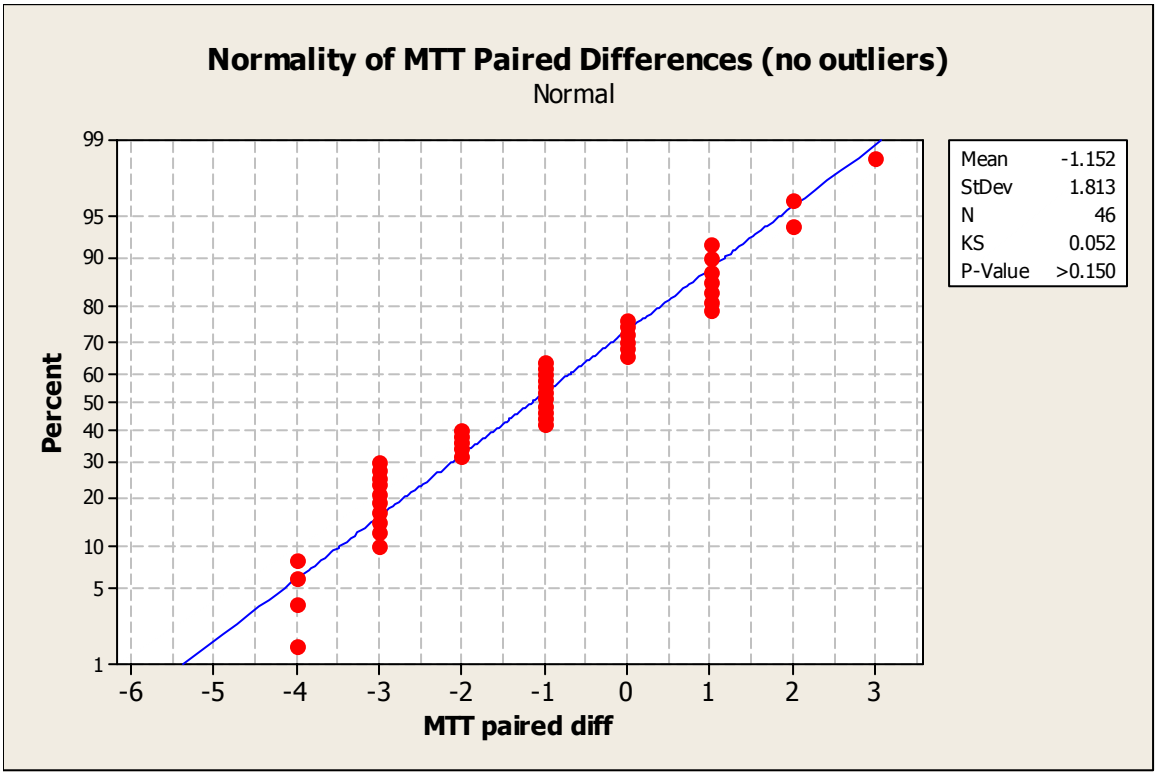


Figure E.20 Normality test for MTT paired differences (no outliers)

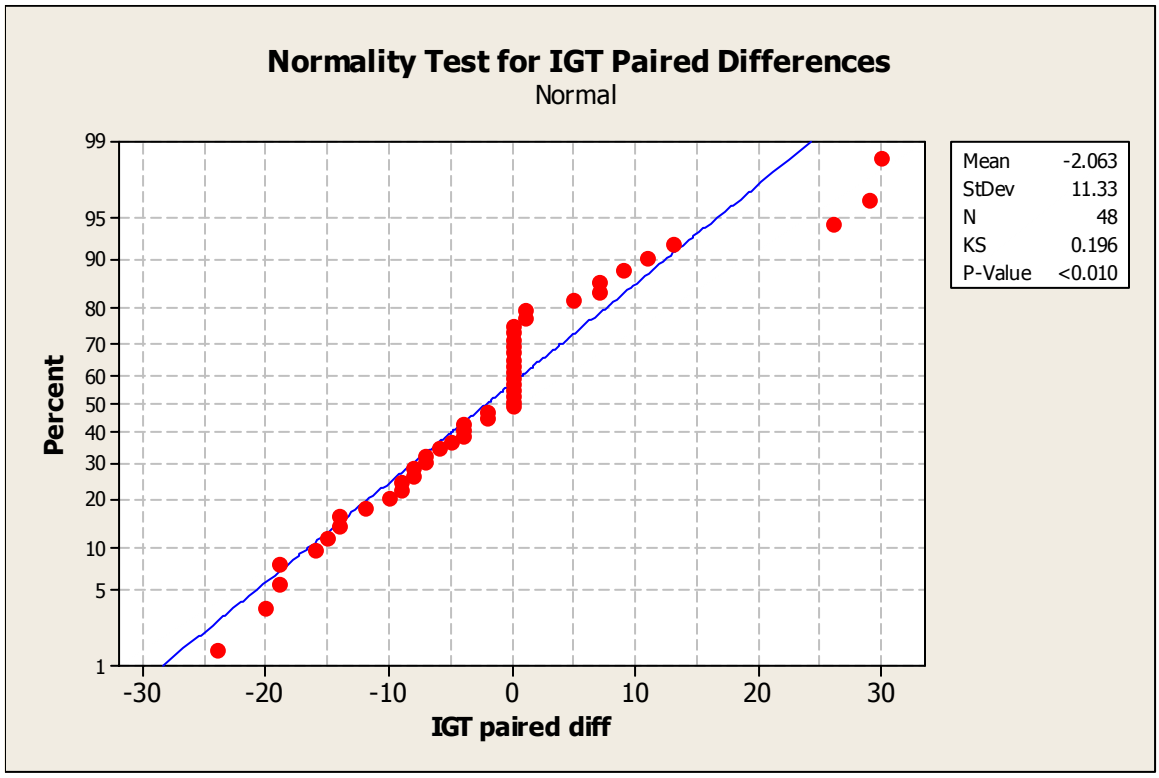


Figure E.21 Normality test for IGT paired differences

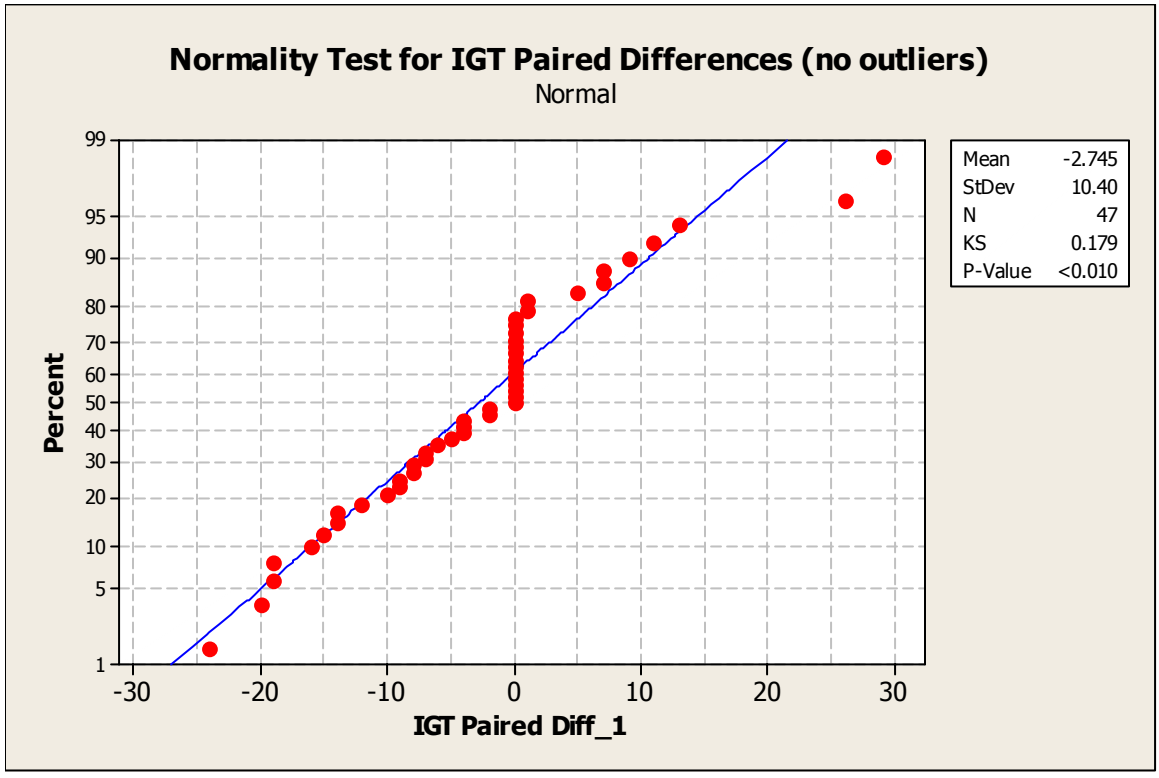


Figure E.22 Normality test for IGT paired differences (no outliers)

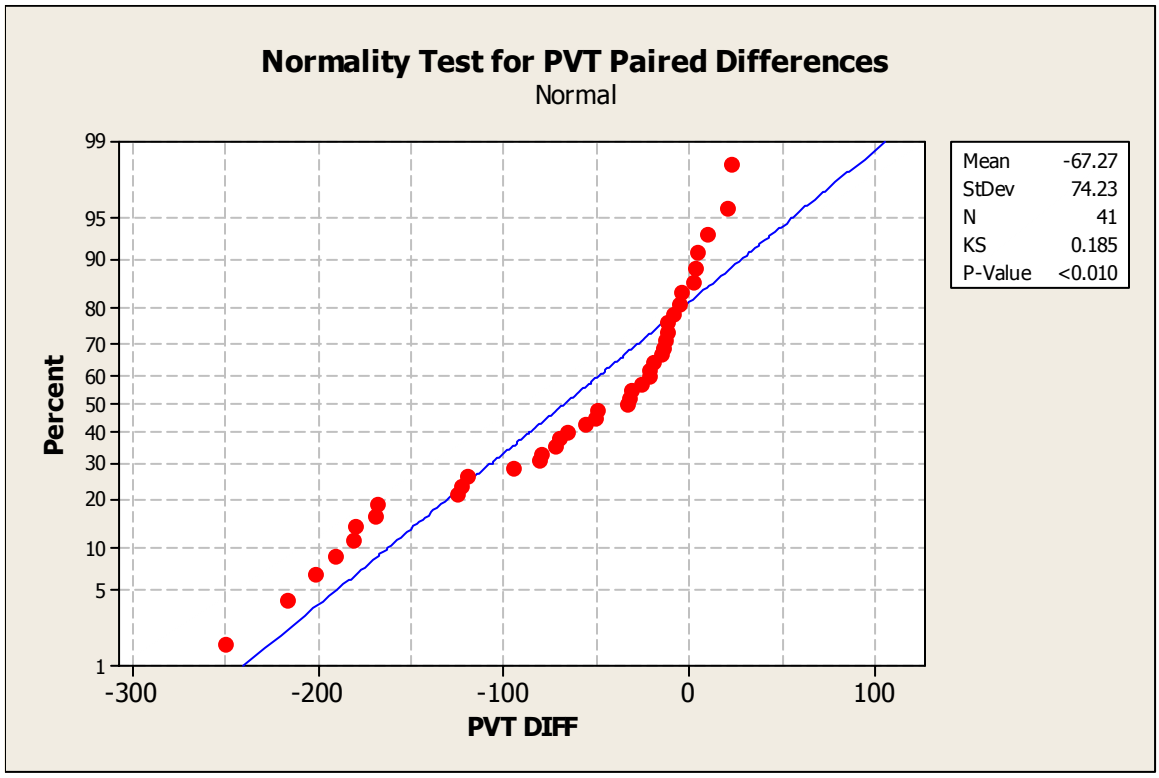


Figure E.23 Normality test for paired differences on PVT

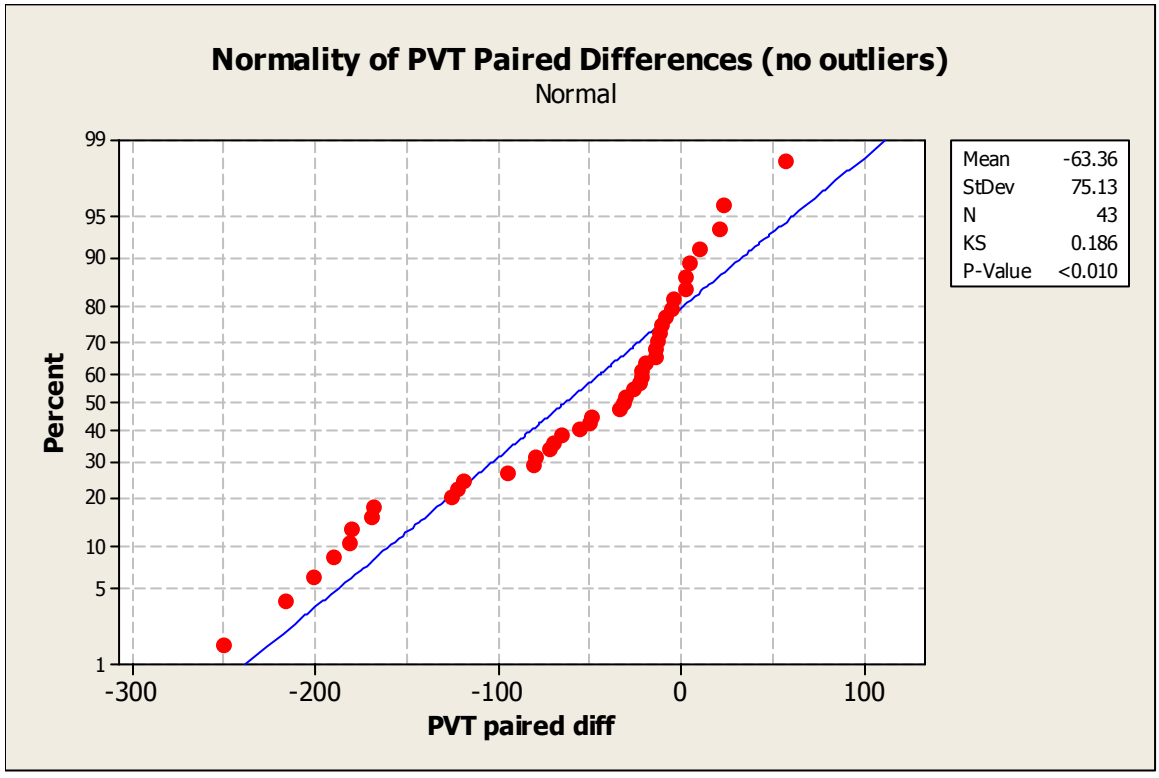


Figure E.24 Normality test for paired differences on PVT (no outliers)

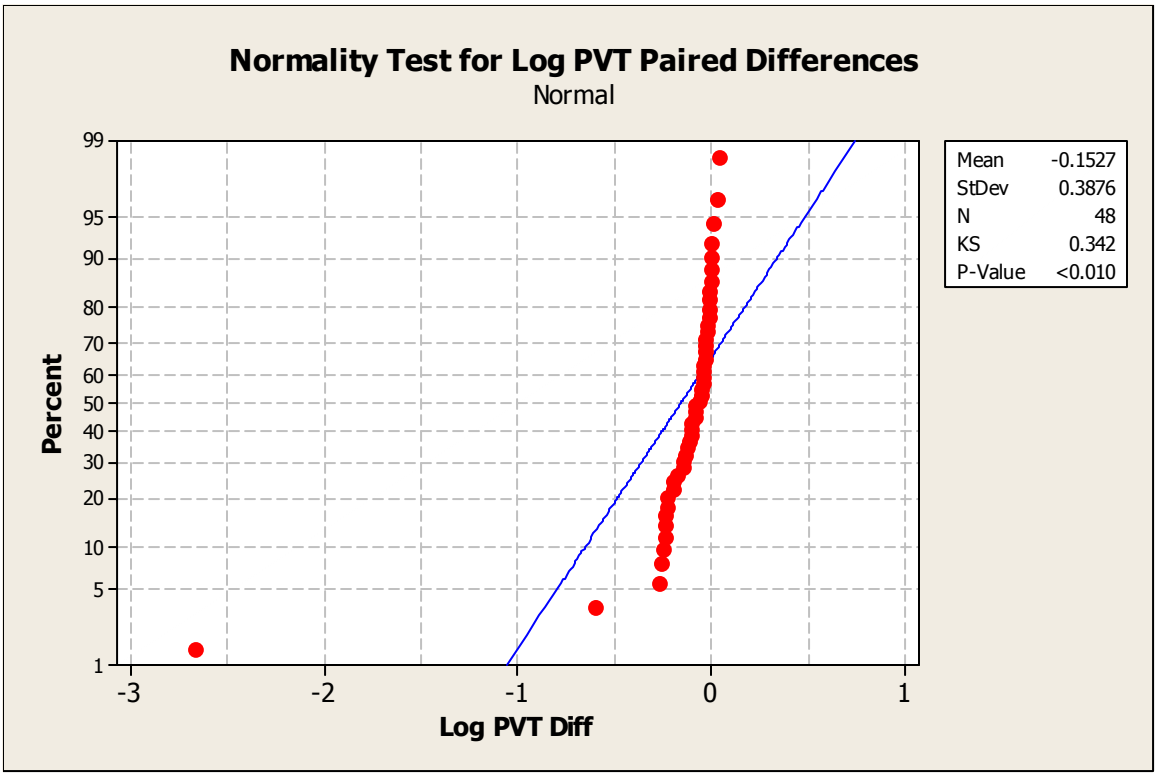


Figure E.25 Normality test for paired differences on Log PVT

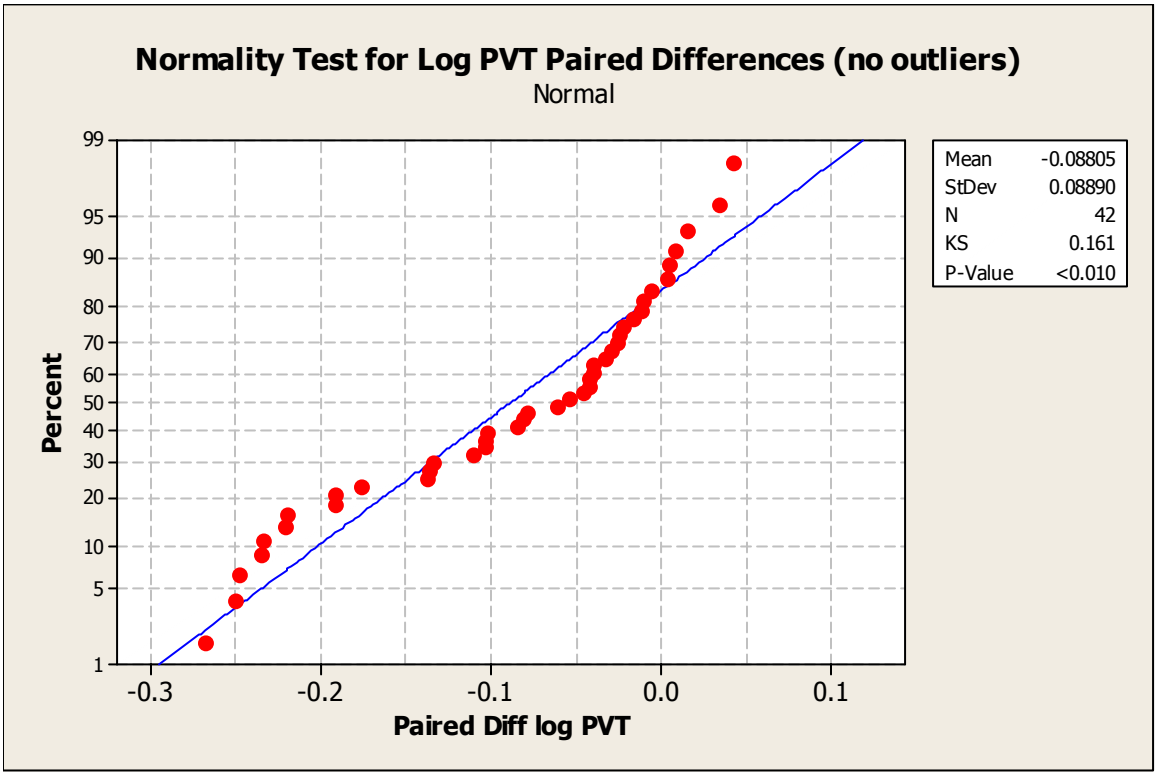


Figure E.26 Normality test for paired differences on Log PVT (no outliers)

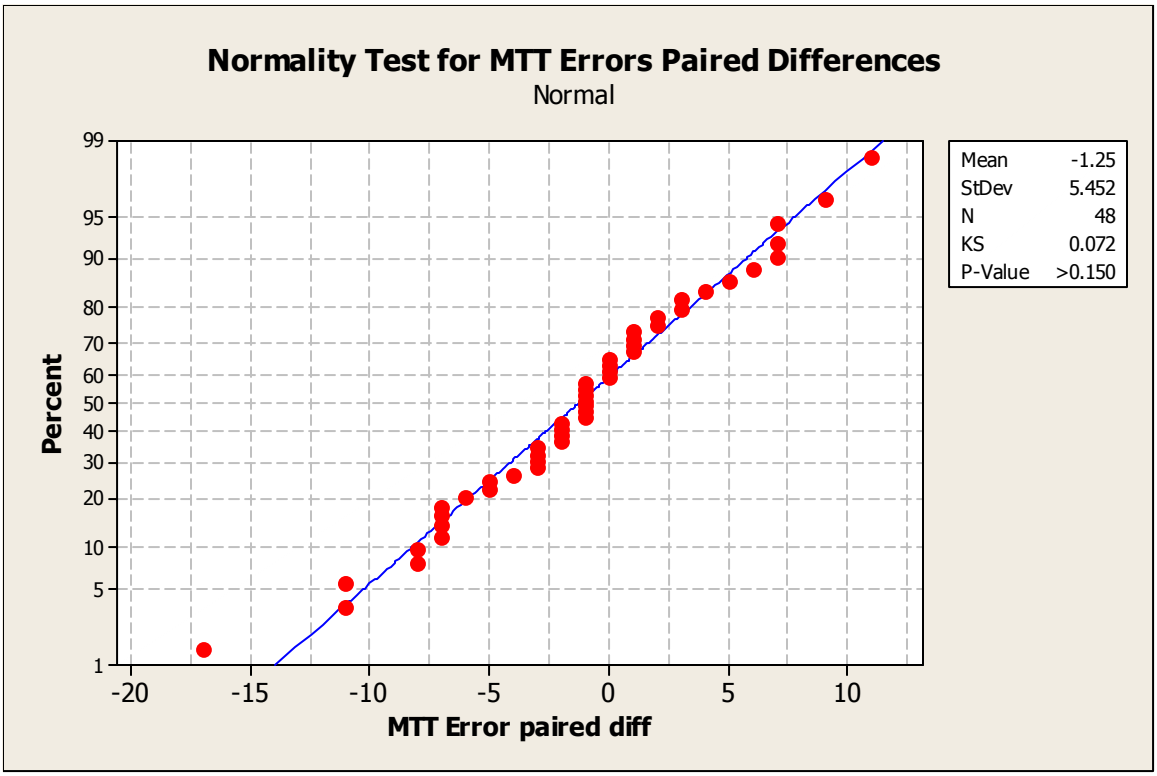


Figure E.27 Normality test for paired differences on MTT Errors

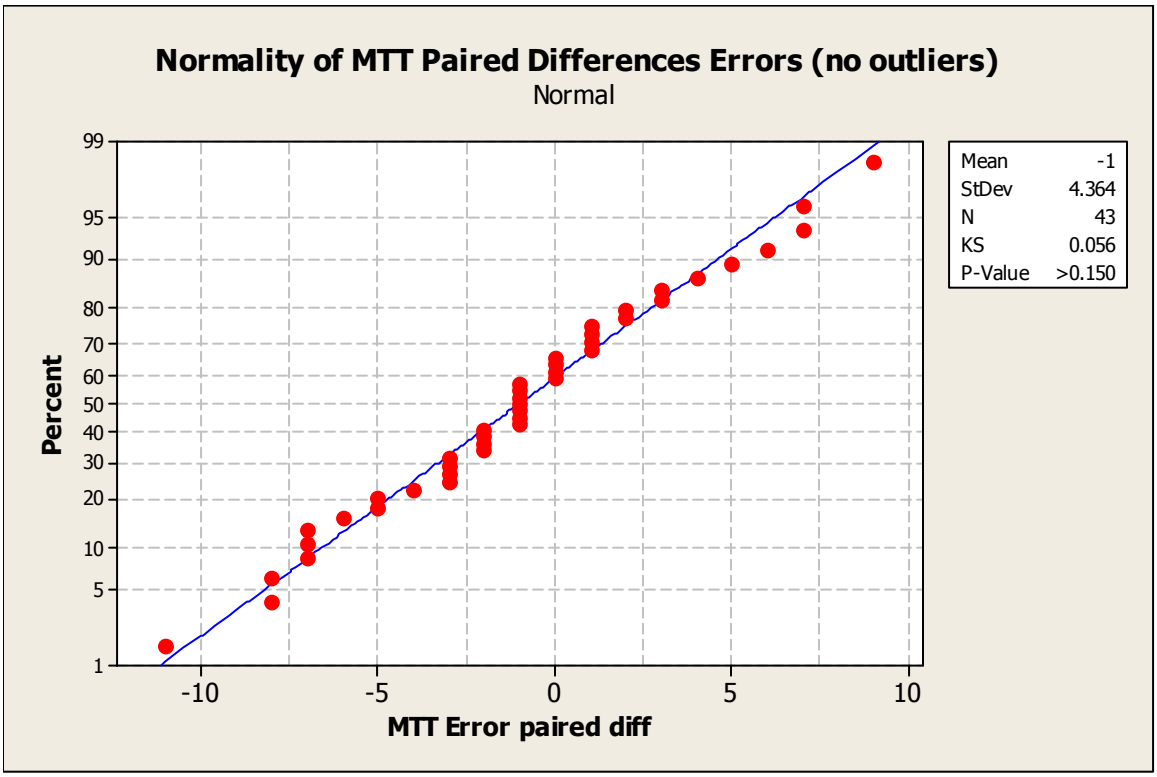


Figure E.28 Normality test for paired differences on MTT Errors (no outliers)

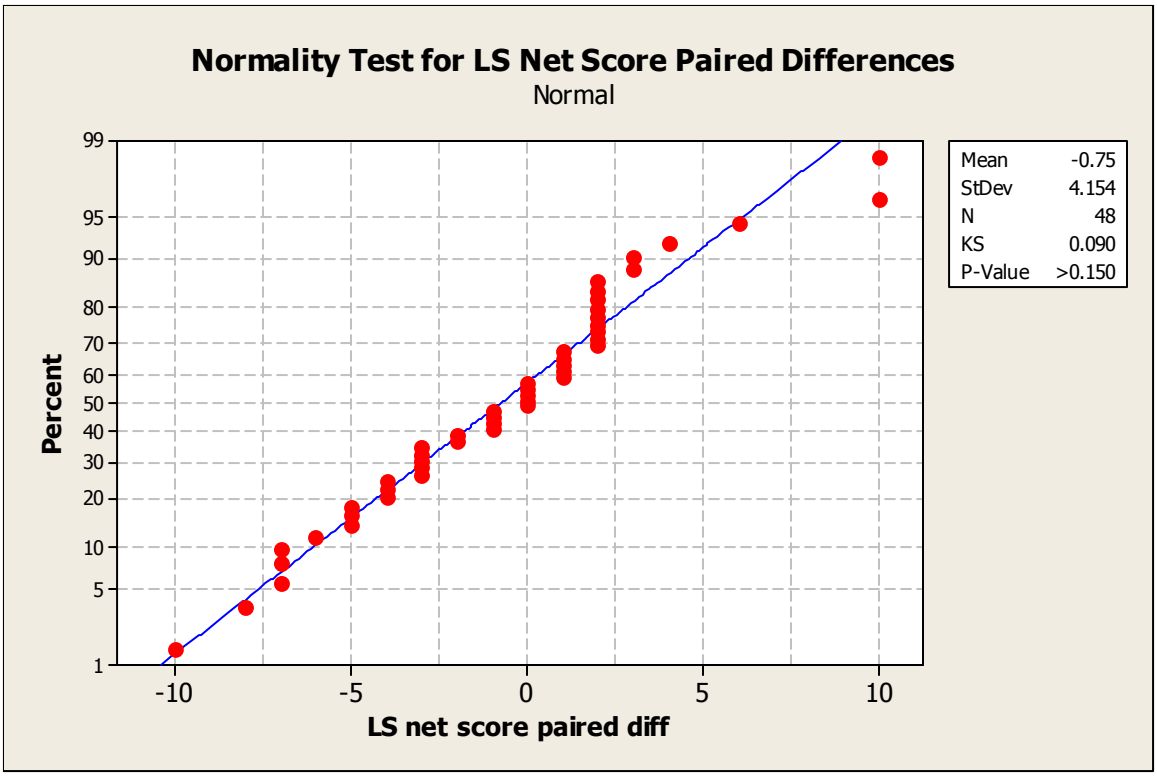


Figure E.29 Normality test for paired differences on LS Net Score

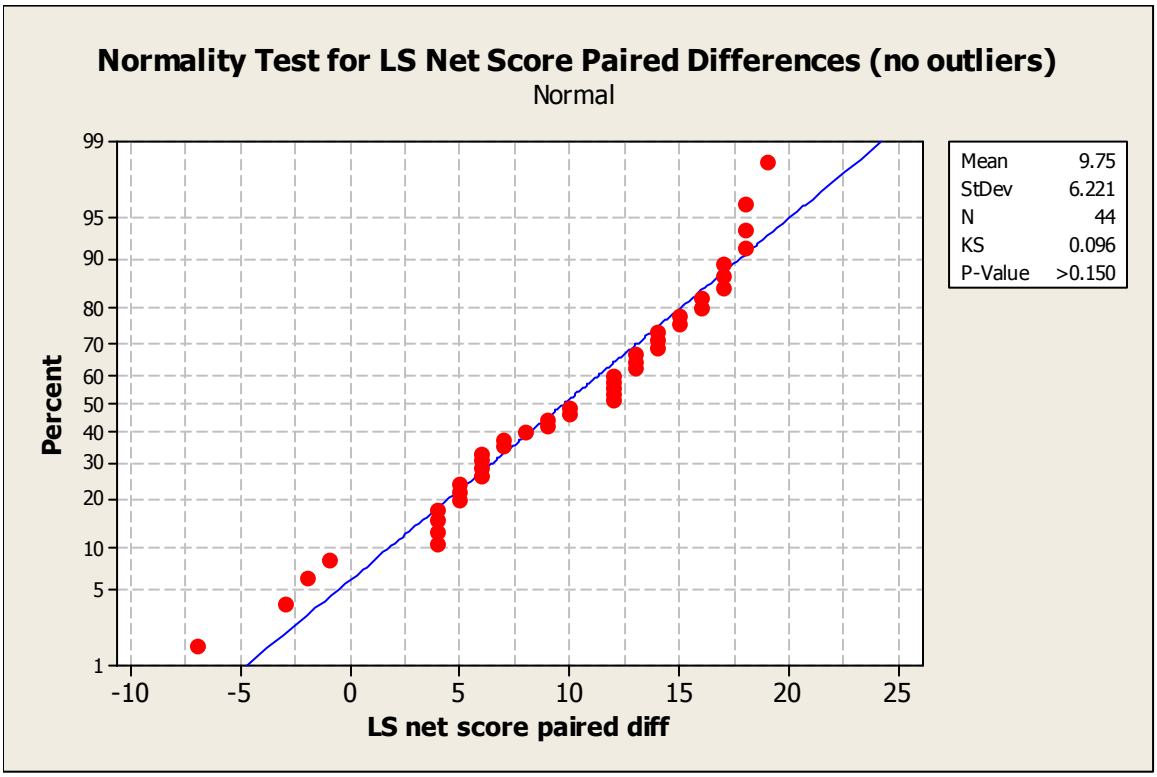


Figure E.30 Normality test for paired differences on LS Net Score (no outliers)

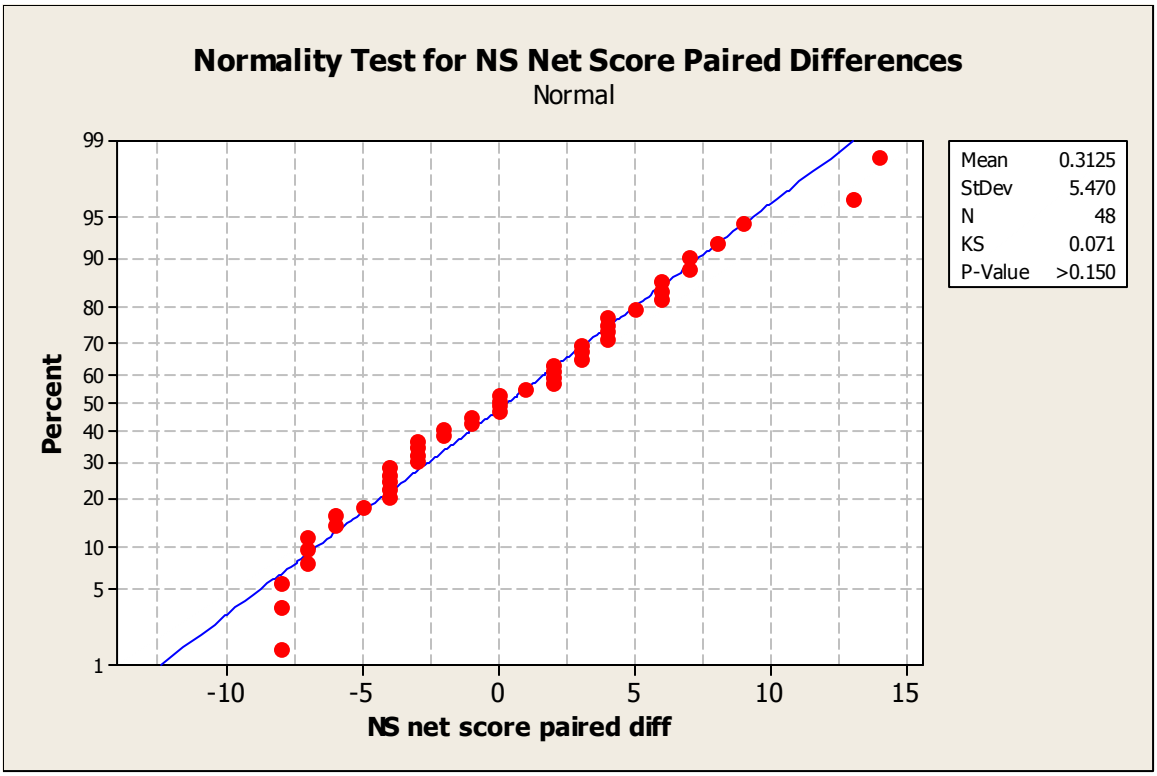


Figure E.31 Normality test for paired differences on NS Net Score

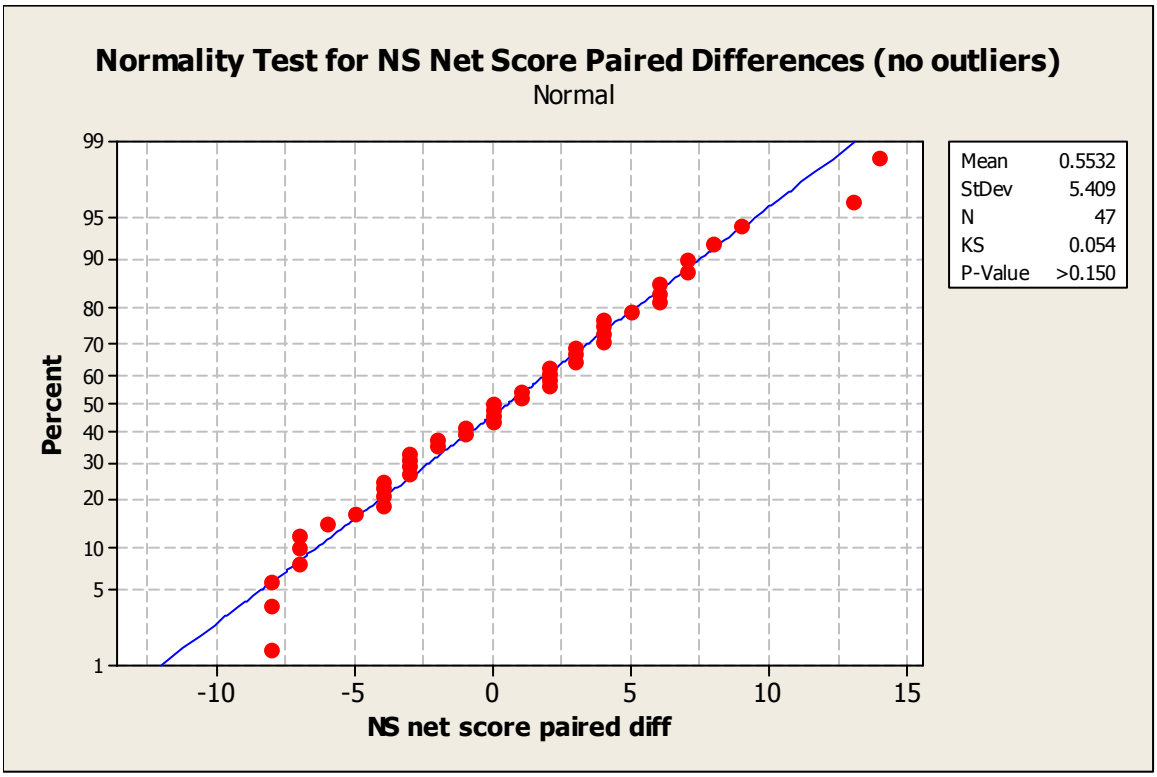


Figure E.32 Normality test for paired differences on NS Net Score (no outliers)

E.2 Paired T-Tests (outliers included)

Paired T-Test and CI: S1 Wombat Overall, S2 WOMBAT Overall

```
Paired T for S1 Wombat Overall - S2 WOMBAT Overall

                    N      Mean     StDev  SE Mean
S1 Wombat Overal   48   94.3375   50.1793   7.2428
S2 WOMBAT Overal   48   98.9333   55.5005   8.0108
Difference         48  -4.59583  27.73238  4.00282


95% CI for mean difference: (-12.64848, 3.45681)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.15  P-Value = 0.257
```

Paired T-Test and CI: S1 LNS-no. correct, S2 LNS-no. correct

```
Paired T for S1 LNS-no. correct - S2 LNS-no. correct

                    N      Mean     StDev   SE Mean
S1 LNS-no. corre   48   8.50000   1.93512   0.27931
S2 LNS-no. corre   48   8.45833   1.80965   0.26120
Difference         48  0.041667  1.890157  0.272821


95% CI for mean difference: (-0.507178, 0.590511)
T-Test of mean difference = 0 (vs not = 0): T-Value = 0.15  P-Value = 0.879
```

Paired T-Test and CI: Baseline LS, SD LS

```
Paired T for Baseline LS - SD LS

               N       Mean     StDev   SE Mean
Baseline LS   48    10.2917    2.7363    0.3950
SD LS         48    10.7292    2.5908    0.3740
Difference    48  -0.437500  2.575715  0.371772


95% CI for mean difference: (-1.185410, 0.310410)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.18  P-Value = 0.245
```

Paired T-Test and CI: Baseline NS, SD NS

```
Paired T for Baseline NS - SD NS

               N      Mean     StDev   SE Mean
Baseline NS   48   7.87500   2.52330   0.36421
SD NS         48   7.39583   3.16053   0.45618
Difference    48  0.479167  3.052656  0.440613


95% CI for mean difference: (-0.407232, 1.365565)
T-Test of mean difference = 0 (vs not = 0): T-Value = 1.09  P-Value = 0.282
```

Paired T-Test and CI: Baseline MTT, SD MTT

```
Paired T for Baseline MTT - SD MTT

               N      Mean    StDev  SE Mean
Baseline MTT  48   12.0417   3.0871   0.4456
SD MTT        48   13.3542   3.6407   0.5255
Difference    48  -1.31250  1.98029  0.28583


95% CI for mean difference: (-1.88752, -0.73748)
T-Test of mean difference = 0 (vs not = 0): T-Value = -4.59  P-Value = 0.000
```

Nonparametric Sign Test: Baseline IGT, SD IGT and Baseline PVT, SD PVT

Frequencies

		N
SDIGT - BaselineIGT	Negative Differences[a,b]	11
	Positive Differences[c,d]	23
	Ties[e,f]	3
	Total	37
PVT14 - PVT5	Negative Differences[a,b]	6
	Positive Differences[c,d]	40
	Ties[e,f]	0
	Total	46

a. SDIGT < BaselineIGT

b. PVT14 < PVT5

c. SDIGT > BaselineIGT

d. PVT14 > PVT5

e. SDIGT = BaselineIGT

f. PVT14 = PVT5

Test Statistics[a]

	SDIGT - BaselineIGT	PVT14 - PVT5
Z	-1.886	-4.866
Asymp. Sig. (2-tailed)	.059	.000

a. Sign Test

Paired T-Test and CI: Baseline IGT, SD IGT

```
Paired T for Baseline IGT - SD IGT

               N      Mean     StDev  SE Mean
Baseline IGT  37   37.8649   14.9744   2.4618
SD IGT        37   40.5405   15.3254   2.5195
Difference    37  -2.67568  12.87559  2.11673


95% CI for mean difference: (-6.96861, 1.61726)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.26  P-Value = 0.214
```

Paired T-Test and CI: 5, 14

```
Paired T for 5 - 14

             N      Mean     StDev  SE Mean
5           46   260.462    46.643    6.877
14          46   344.370   135.920   20.040
Difference  46  -83.9083  117.6447  17.3458


95% CI for mean difference: (-118.8444, -48.9721)
T-Test of mean difference = 0 (vs not = 0): T-Value = -4.84  P-Value = 0.000
```

Paired T-Test and CI: Base LS_crct - incrt, SD LS_crct - incrt

```
Paired T for Base LS_crct - incrt - SD LS_crct - incrt

                   N       Mean     StDev   SE Mean
Base LS_crct - i  48    8.56250   3.97281   0.57343
SD LS_crct - inc  48    9.31250   4.10581   0.59262
Difference        48  -0.750000  4.153952  0.599571


95% CI for mean difference: (-1.956182, 0.456182)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.25  P-Value = 0.217
```

Paired T-Test and CI: Base NS_crct - incrt, SD NS_crct - incrt

```
Paired T for Base NS_crct - incrt - SD NS_crct - incrt

                   N      Mean     StDev   SE Mean
Base NS_crct - i  48   2.70833   4.15118   0.59917
SD NS_crct - inc  48   2.39583   5.11460   0.73823
Difference        48  0.312500  5.470059  0.789535


95% CI for mean difference: (-1.275839, 1.900839)
T-Test of mean difference = 0 (vs not = 0): T-Value = 0.40  P-Value = 0.694
```

Paired T-Test and CI: Baseline MTT Errors, SD MTT Errors

```
Paired T for Baseline MTT Errors - SD MTT Errors

                    N      Mean    StDev  SE Mean
Baseline MTT Err   48   4.41667  3.32517  0.47995
SD MTT Errors      48   5.66667  4.11191  0.59350
Difference         48  -1.25000  5.45192  0.78692


95% CI for mean difference: (-2.83307, 0.33307)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.59  P-Value = 0.119
```

E.3 Paired T-Tests (no outliers)

Paired T-Test and CI: Baseline LNS, SD LNS

```
Paired T for Baseline LNS - SD LNS

                N       Mean     StDev   SE Mean
Baseline LNS   46    8.26087   1.56964   0.23143
SD LNS         46    8.36957   1.76835   0.26073
Difference     46  -0.108696  1.779242  0.262335


95% CI for mean difference: (-0.637065, 0.419674)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.41  P-Value = 0.681
```

Paired T-Test and CI: Baseline LS, SD LS

```
Paired T for Baseline LS - SD LS

               N       Mean     StDev   SE Mean
Baseline LS   47    10.4468    2.5436    0.3710
SD LS         47    10.7660    2.6061    0.3801
Difference    47  -0.319149  2.468113  0.360011


95% CI for mean difference: (-1.043814, 0.405516)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.89  P-Value = 0.380
```

Paired T-Test and CI: Baseline NS, SD NS

```
Paired T for Baseline NS - SD NS

               N      Mean     StDev   SE Mean
Baseline NS   45   7.55556   2.26189   0.33718
SD NS         45   7.11111   3.02431   0.45084
Difference    45  0.444444  3.100994  0.462269


95% CI for mean difference: (-0.487197, 1.376086)
T-Test of mean difference = 0 (vs not = 0): T-Value = 0.96  P-Value = 0.342
```

Paired T-Test and CI: Baseline MTT, SD MTT

```
Paired T for Baseline MTT - SD MTT

               N      Mean    StDev  SE Mean
Baseline MTT  45   11.6889   2.8028   0.4178
SD MTT        45   12.9333   3.0406   0.4533
Difference    45  -1.24444  1.72093  0.25654


95% CI for mean difference: (-1.76147, -0.72742)
T-Test of mean difference = 0 (vs not = 0): T-Value = -4.85  P-Value = 0.000
```

Nonparametric Sign Test for Baseline IGT, SD IGT and Baseline PVT, SD PVT

Frequencies

		N
SD IGT - Baseline IGT	Negative Differences[a,b]	11
	Positive Differences[c,d]	21
	Ties[e,f]	3
	Total	35
PVT 14 - PVT 5	Negative Differences[a,b]	7
	Positive Differences[c,d]	36
	Ties[e,f]	0
	Total	43

a. SD IGT < Baseline IGT

b. PVT 14 < PVT 5

c. SD IGT > Baseline IGT

d. PVT 14 > PVT 5

e. SD IGT = Baseline IGT

f. PVT 14 = PVT 5

Test Statistics[a]

	SD IGT - Baseline IGT	PVT 14 - PVT 5
Z	-1.591	-4.270
Asymp. Sig. (2-tailed)	.112	.000

a. Sign Test

Paired T-Test and CI: Baseline LS, SD LS

```
Paired T for Baseline LS - SD LS

              N       Mean     StDev   SE Mean
Baseline LS  47    10.4468    2.5436    0.3710
SD LS        47    10.7660    2.6061    0.3801
Difference   47  -0.319149  2.468113  0.360011


95% CI for mean difference: (-1.043814, 0.405516)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.89  P-Value = 0.380
```

Paired T-Test and CI: Baseline IGT, SD IGT

```
Paired T for Baseline IGT - SD IGT

               N      Mean     StDev  SE Mean
Baseline IGT  34   36.7647   13.7732   2.3621
SD IGT        34   39.9706   15.8716   2.7220
Difference    34  -3.20588  11.94883  2.04921


95% CI for mean difference: (-7.37503, 0.96326)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.56  P-Value = 0.127
```

Paired T-Test and CI: PVT 5, PVT 14

```
Paired T for PVT 5 - PVT 14

             N      Mean    StDev  SE Mean
PVT 5       43   256.186   39.044    5.954
PVT 14      43   319.543   93.072   14.193
Difference  43  -63.3565  75.1320  11.4575


95% CI for mean difference: (-86.4787, -40.2343)
T-Test of mean difference = 0 (vs not = 0): T-Value = -5.53  P-Value = 0.000
```

Paired T-Test and CI: Base LS_crct - incrt, SD LS_crct - incrt

```
Paired T for Base LS_crct - incrt - SD LS_crct - incrt

                  N       Mean     StDev   SE Mean
Base LS_crct - i  44     9.1818    3.3499    0.5050
SD LS_crct - inc  44    10.1136    2.5538    0.3850
Difference        44  -0.931818  3.322914  0.500948


95% CI for mean difference: (-1.942076, 0.078440)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.86  P-Value = 0.070
```

Paired T-Test and CI: Base NS_crct - incrt, SD NS_crct - incrt

```
Paired T for Base NS_crct - incrt - SD NS_crct - incrt
```

```
                    N      Mean     StDev   SE Mean
Base NS_crct - i   46   2.78261   3.87535   0.57139
SD NS_crct - inc   46   2.23913   4.95394   0.73042
Difference         46  0.543478  5.467913  0.806200


95% CI for mean difference: (-1.080291, 2.167248)
T-Test of mean difference = 0 (vs not = 0): T-Value = 0.67  P-Value = 0.504
```

Paired T-Test and CI: Baseline MTT Errors, SD MTT Errors

```
Paired T for Baseline MTT Errors - SD MTT Errors

                    N      Mean    StDev  SE Mean
Baseline MTT Err   43   4.02326  2.93167  0.44708
SD MTT Errors      43   5.02326  3.13573  0.47819
Difference         43  -1.00000  4.36436  0.66556


95% CI for mean difference: (-2.34315, 0.34315)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.50  P-Value = 0.140
```

APPENDIX F. POWER ANALYSIS

F.1 Power Analysis with outliers in the data

F.1.1 Power for a test of the null hypothesis—WOMBAT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 20.3% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 4.6 with a standard deviation of 27.7. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.2 Power for a test of the null hypothesis—LNS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 5.0% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.0 with a standard deviation of 1.9. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.3 Power for a test of the null hypothesis—LS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 18.1% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.4 with a standard deviation of 2.6. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is

reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.4 Power for a test of the null hypothesis—NS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 19.5% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.5 with a standard deviation of 3.1. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.5 Power for a test of the null hypothesis—MTT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 99.3% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 1.3 with a standard deviation of 2.0. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.6 Power for a test of the null hypothesis—IGT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 37 pairs of cases, the study will have power of 24.3% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 2.1 with a standard deviation of 11.3. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.1.7 Power for a test of the null hypothesis—PVT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 99.8% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 90.0 with a standard deviation of 128.0. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2 Power Analysis (no outliers)

F.2.1 Power for a test of the null hypothesis—WOMBAT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 21.4% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 4.7 with a standard deviation of 27.4. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.2 Power for a test of the null hypothesis—LNS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 46 pairs of cases, the study will have power of 6.6% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.1 with a standard deviation of 1.8. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.3 Power for a test of the null hypothesis—LS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 46 pairs of cases, the study will have power of 49.7% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.5 with a standard deviation of 1.7. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.4 Power for a test of the null hypothesis—NS

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 48 pairs of cases, the study will have power of 11.7% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 0.3 with a standard deviation of 2.7. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.5 Power for a test of the null hypothesis—MTT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within

pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 46 pairs of cases, the study will have power of 82.3% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 1.0 with a standard deviation of 2.3. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.6 Power for a test of the null hypothesis—IGT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 37 pairs of cases, the study will have power of 43.9% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 3.6 with a standard deviation of 11.8. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

F.2.7 Power for a test of the null hypothesis—PVT

One goal of the proposed study is to test the null hypothesis that the mean difference (or change) within pairs is 0.00. The criterion for significance (alpha) has been set at 0.050. The test is 2-tailed, which means that an effect in either direction will be interpreted.

With the proposed sample size of 43 pairs of cases, the study will have power of exceeding 99.9% to yield a statistically significant result.

This computation assumes that the population from which the sample will be drawn has a mean difference of 64.8 with a standard deviation of 73.4. The observed value will be tested against a theoretical value (constant) of 0.00

This effect was selected as the smallest effect that would be important to detect, in the sense that any smaller effect would not be of clinical or substantive significance. It is also assumed that this effect size is reasonable, in the sense that an effect of this magnitude could be anticipated in this field of research.

APPENDIX G. MULTIPLE LINEAR REGRESSION (WOMBAT)

G.1 Variable Selection with Outliers in Data

G.1.1 Baseline Variable Selection

G.1.1.1 Backward Elimination:

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is Baseline WOMBAT on 6 predictors, with N = 35
N(cases with missing observations) = 13 N(all cases) = 48


Step                 1        2        3        4
Constant        -90.82   -89.50  -115.05  -146.57

Baseline LNS       7.5      7.6      7.7      8.4
T-Value           1.86     2.01     2.05     2.27
P-Value          0.073    0.054    0.049    0.030

Baseline LS        9.7      9.7     10.0     10.5
T-Value           3.78     3.86     4.14     4.32
P-Value          0.001    0.001    0.000    0.000

Baseline NS        0.2
T-Value           0.05
P-Value          0.958

Baseline MTT       4.9      4.9      5.1      5.1
T-Value           2.07     2.11     2.19     2.18
P-Value          0.048    0.044    0.037    0.037

Baseline IGT     -0.52    -0.52    -0.55
T-Value          -1.12    -1.14    -1.23
P-Value          0.274    0.263    0.228

5                -0.08    -0.08
T-Value          -0.52    -0.54
P-Value          0.610    0.594

S                 39.8     39.1     38.6     38.9
R-Sq             54.31    54.30    53.84    51.51
R-Sq(adj)        44.51    46.42    47.69    46.82
Mallows C-p        7.0      5.0      3.3      2.7
PRESS          63445.0  58432.6  57333.0  57908.1
R-Sq(pred)       34.59    39.76    40.89    40.30
```

G.1.1.2 Best Subsets Regression:

```
Response is Baseline WOMBAT

35 cases used, 13 cases contain missing values

                                        B     B B
```

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	a s e l i n e L N	B a s e l i n e L S	B a s e l i n e S S	a s e l i n e M T S	a s e l i n e I G T	a s e l i n e 5
1	33.5	31.4	9.8	44.226		X				
1	15.6	13.1	20.7	49.792	X					
2	44.1	40.6	5.3	41.168	X	X				
2	43.4	39.9	5.7	41.404		X		X		
3	51.5	46.8	2.7	38.950	X	X		X		
3	47.4	42.3	5.3	40.582		X		X	X	
4	53.8	47.7	3.3	38.631	X	X		X	X	
4	52.2	45.9	4.3	39.292	X	X		X		X
5	54.3	46.4	5.0	39.095	X	X		X	X	X
5	53.9	45.9	5.3	39.279	X	X	X	X	X	
6	54.3	44.5	7.0	39.785	X	X	X	X	X	X

G.1.2 Sleep Deprived Variable Selection

G.1.2.1 Backward Elimination.

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is SD WOMBAT on 6 predictors, with N = 36
N(cases with missing observations) = 12 N(all cases) = 48


Step              1       2        3        4
Constant     -4.8450  1.1281  -0.7702  17.2449

SD LNS           3.3     3.4      2.9
T-Value         0.75    0.78     0.72
P-Value        0.457   0.441    0.474

SD LS            6.2     6.0      6.0      6.9
T-Value         1.82    1.83     1.84     2.31
P-Value        0.079   0.078    0.076    0.027

SD NS           -0.8    -0.8
T-Value        -0.33   -0.33
P-Value        0.746   0.747

SD MTT           5.2     5.1      5.0      5.0
T-Value         2.28    2.30     2.31     2.33
P-Value        0.030   0.029    0.028    0.026

SD IGT         -1.51   -1.48    -1.44    -1.51
T-Value        -2.93   -2.99    -3.05    -3.30
P-Value        0.007   0.006    0.005    0.002
```

```
14             0.015
T-Value         0.29
P-Value        0.774

S               41.4    40.7     40.1     39.8
R-Sq           52.87   52.74    52.57    51.77
R-Sq(adj)      43.12   44.86    46.45    47.24
Mallows C-p      7.0     5.1      3.2      1.7
```

G.1.2.2 Best Subsets Regression:

```
Response is SD WOMBAT

36 cases used, 12 cases contain missing values
```

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	SD LN	S S L	S D N	S D T	S D I M	S D I G	1 4
1	30.2	28.2	10.9	46.486						X	
1	23.9	21.7	14.8	48.548					X		
2	43.7	40.3	4.6	42.383					X	X	
2	43.6	40.2	4.7	42.432		X				X	
3	51.8	47.2	1.7	39.840		X			X	X	
3	47.4	42.5	4.4	41.603	X				X	X	
4	52.6	46.4	3.2	40.139	X	X			X	X	
4	51.9	45.7	3.6	40.406		X			X	X	X
5	52.7	44.9	5.1	40.731	X	X	X	X	X		
5	52.7	44.8	5.1	40.746	X	X			X	X	X
6	52.9	43.1	7.0	41.367	X	X	X	X	X	X	

G.1.3 Baseline Regression Plots

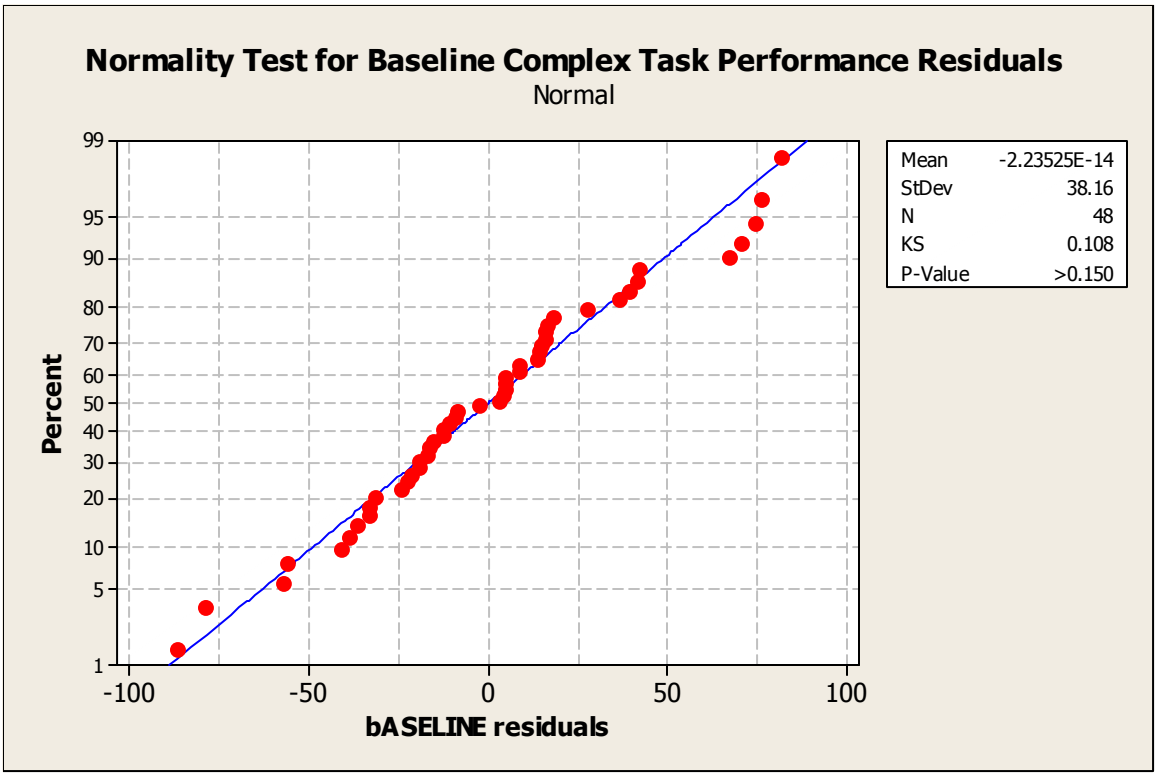


Figure G.1.3.1 Normality Test for Baseline Residuals

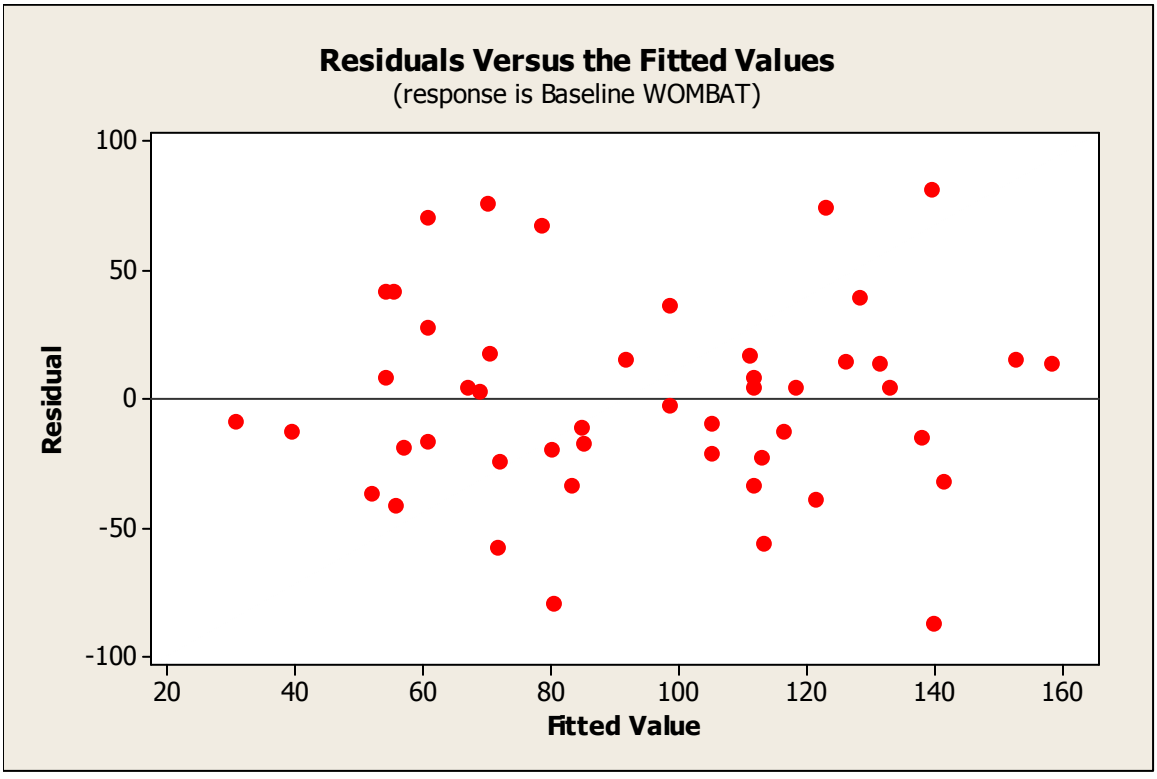


Figure G.1.3.2 Resdiduals vs. Fitted Values (Baseline)

G.1.4 Sleep Deprived Regression Plots

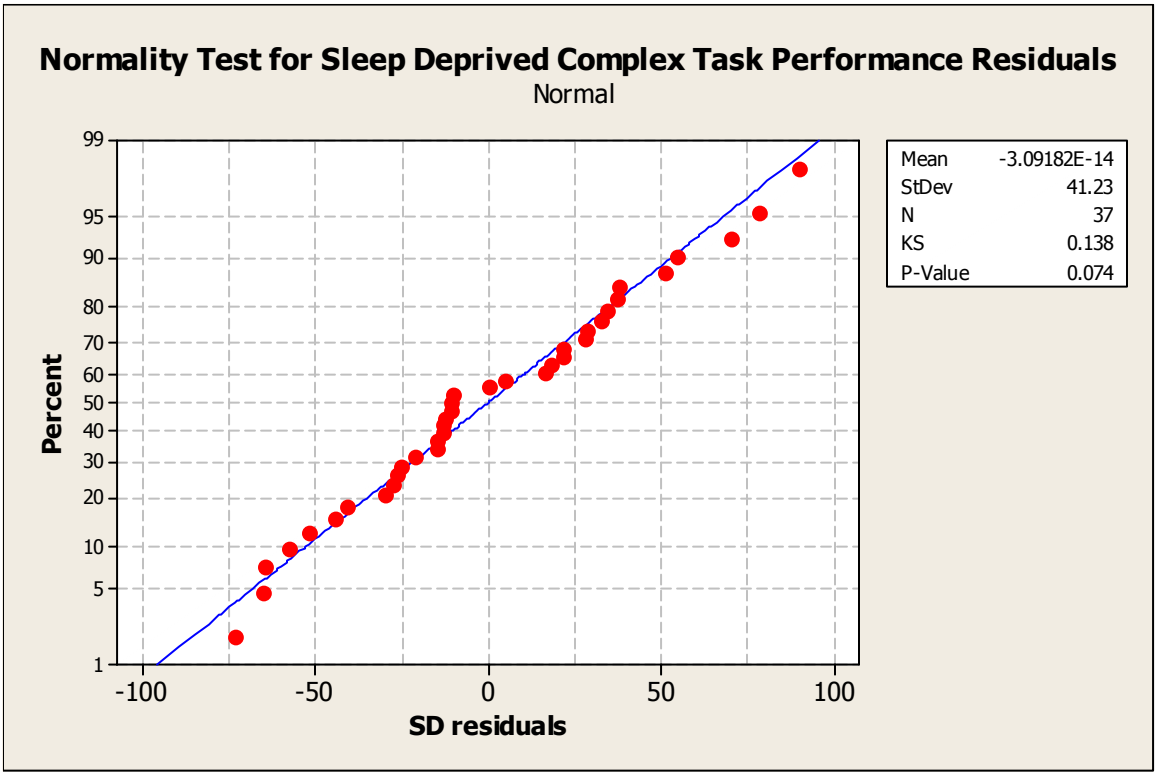


Figure G.1.4.1 Normality Test for Sleep Deprived Residuals

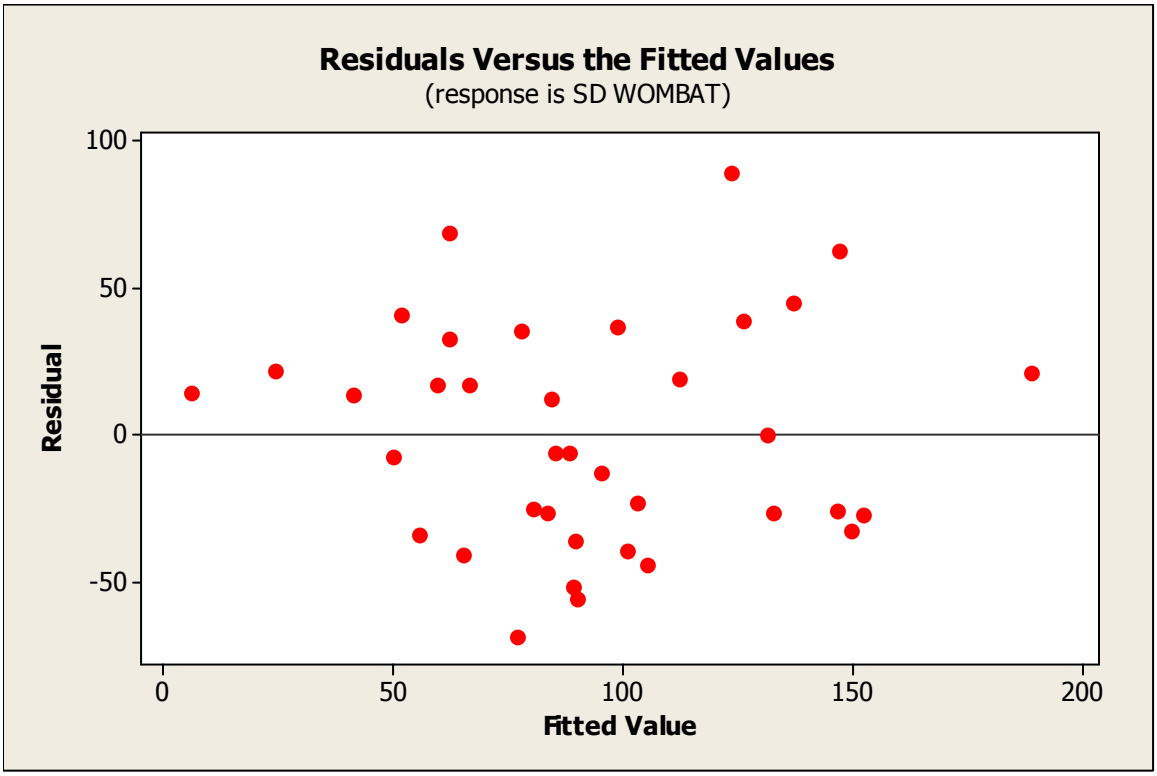


Figure G.1.4.2 Resdiduals vs. Fitted Values (Sleep Deprived)

G.2 Variable Selection (No Outliers in Data)

G.2.1 Baseline Variable Selection

G.2.1.1 Backward Elimination

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is Baseline WOMBAT on 6 predictors, with N = 32
N(cases with missing observations) = 16 N(all cases) = 48


Step               1        2        3        4
Constant      -206.3   -204.0   -211.3   -146.3

SD LNS          -2.8
T-Value        -0.51
P-Value        0.614

SD LS            8.5      7.7      8.0      7.0
T-Value         2.04     2.02     2.15     1.97
P-Value        0.052    0.054    0.041    0.058

SD NS            2.1      1.6
T-Value         0.57     0.47
P-Value        0.571    0.645

SD MTT           7.0      6.7      6.9      6.2
T-Value         2.17     2.15     2.28     2.13
P-Value        0.039    0.041    0.031    0.042

SD IGT          1.46     1.37     1.58     1.31
T-Value         1.71     1.66     2.33     2.16
P-Value        0.100    0.108    0.027    0.039

PVT 14         0.102    0.089    0.088
T-Value         0.99     0.90     0.90
P-Value        0.332    0.374    0.375

S               46.5     45.8     45.1     45.0
R-Sq           41.30    40.69    40.19    38.39
R-Sq(adj)      27.21    29.28    31.33    31.79
Mallows C-p      7.0      5.3      3.5      2.2
PRESS        85384.7  79123.9  75365.2  72261.6
R-Sq(pred)      7.12    13.93    18.02    21.40
```

G.2.1.2 Best Subsets Regression

```
Response is Baseline WOMBAT

29 cases used, 19 cases contain missing values
```

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	Baseline SL	BaseLine NL	BaseLine SS	BaseLine NE	BaseLine MT	BaseLine IGT	BaseLine PVT
1	28.7	26.1	6.0	46.402		X					
1	15.5	12.4	11.8	50.512					X		
2	41.5	37.0	2.5	42.825		X			X		
2	36.8	31.9	4.5	44.531	X	X					
3	46.9	40.5	2.1	41.628	X	X			X		
3	43.4	36.6	3.7	42.982		X			X	X	
4	48.7	40.2	3.3	41.751	X	X			X	X	
4	47.6	38.9	3.8	42.207	X	X	X	X			
5	49.3	38.2	5.1	42.419	X	X	X	X	X		
5	49.0	38.0	5.2	42.513	X	X			X	X	X
6	49.5	35.7	7.0	43.275	X	X	X	X	X	X	

G.2.2 Sleep Deprived Variable Selection

G.2.2.1 Backward Elimination

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is SD WOMBAT on 6 predictors, with N = 32
N(cases with missing observations) = 16 N(all cases) = 48


Step             1       2      3       4       5
Constant    14.182  21.329  7.279  36.342  87.288

SD LNS         2.5
T-Value       0.49
P-Value      0.625

SD LS          5.5     6.1    6.1     5.5
T-Value       1.47    1.75   1.78    1.67
P-Value      0.154   0.092  0.086   0.106

SD NS         -1.8    -1.3
T-Value      -0.56   -0.43
P-Value      0.580   0.668

SD MTT         5.2     5.5    5.3     4.8     5.9
T-Value       1.75    1.91   1.90    1.78    2.22
P-Value      0.092   0.067  0.068   0.085   0.035

SD IGT       -1.75   -1.81  -1.70   -1.61   -1.77
T-Value      -2.93   -3.13  -3.30   -3.25   -3.55
P-Value      0.007   0.004  0.003   0.003   0.001
```

```
PVT 14         0.045   0.054  0.062
T-Value         0.50    0.63   0.74
P-Value        0.622   0.536  0.464

S               43.0    42.4   41.8    41.4    42.7
R-Sq           48.14   47.63  47.25   46.17   40.80
R-Sq(adj)      35.69   37.56  39.44   40.40   36.72
Mallows C-p      7.0     5.2    3.4     1.9     2.5
```

G.2.2.2 Best Subsets Regression

```
Response is SD WOMBAT

32 cases used, 16 cases contain missing values
```

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	SD LN	S SD LS	S SD NS	S SD LT	S SD IG	PVT 14
1	30.8	28.5	5.4	45.393					X	
1	17.9	15.2	11.6	49.437		X				
2	40.8	36.7	2.5	42.696				X	X	
2	40.1	35.9	2.9	42.961		X			X	
3	46.2	40.4	1.9	41.434		X		X	X	
3	42.4	36.2	3.8	42.867	X			X	X	
4	47.3	39.4	3.4	41.769		X		X	X	X
4	46.8	39.0	3.6	41.932		X	X	X	X	
5	47.6	37.6	5.2	42.411		X	X	X	X	X
5	47.6	37.5	5.2	42.415	X	X	X	X	X	
6	48.1	35.7	7.0	43.041	X	X	X	X	X	X

G.2.3 Baseline Regression Plots

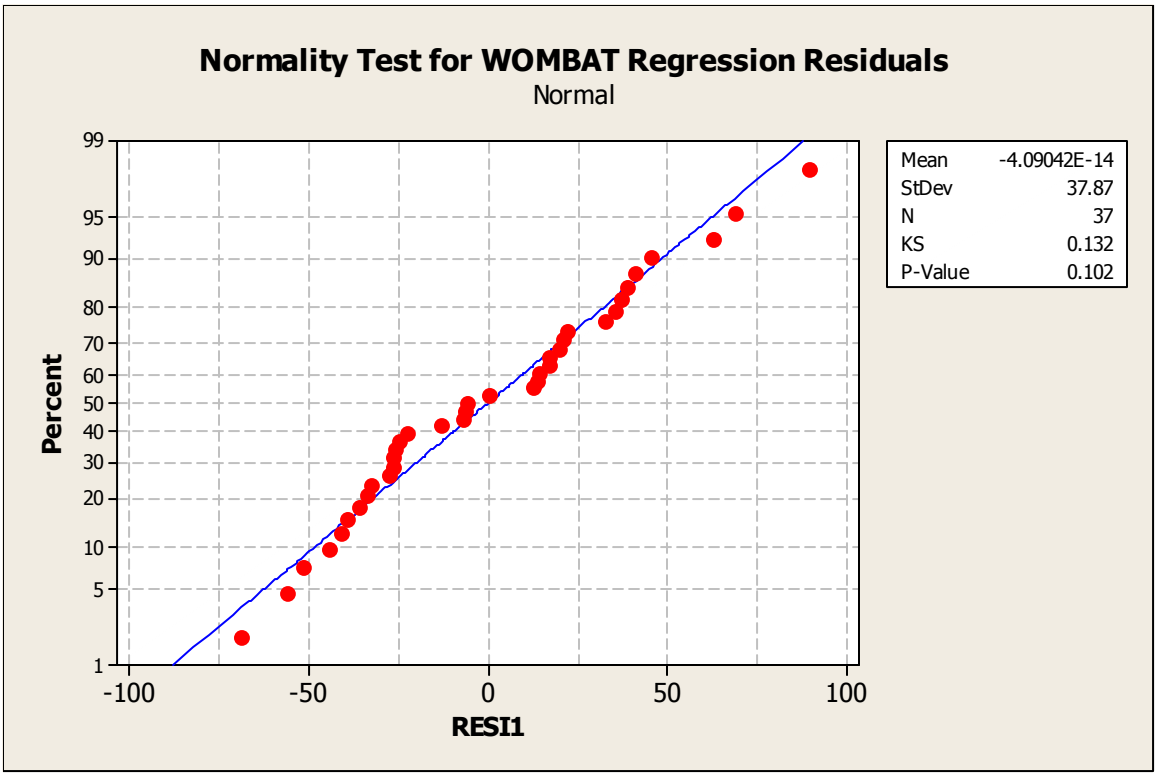


Figure G.2.3.1 Normailty Test for Baseline Residuals (no outliers)

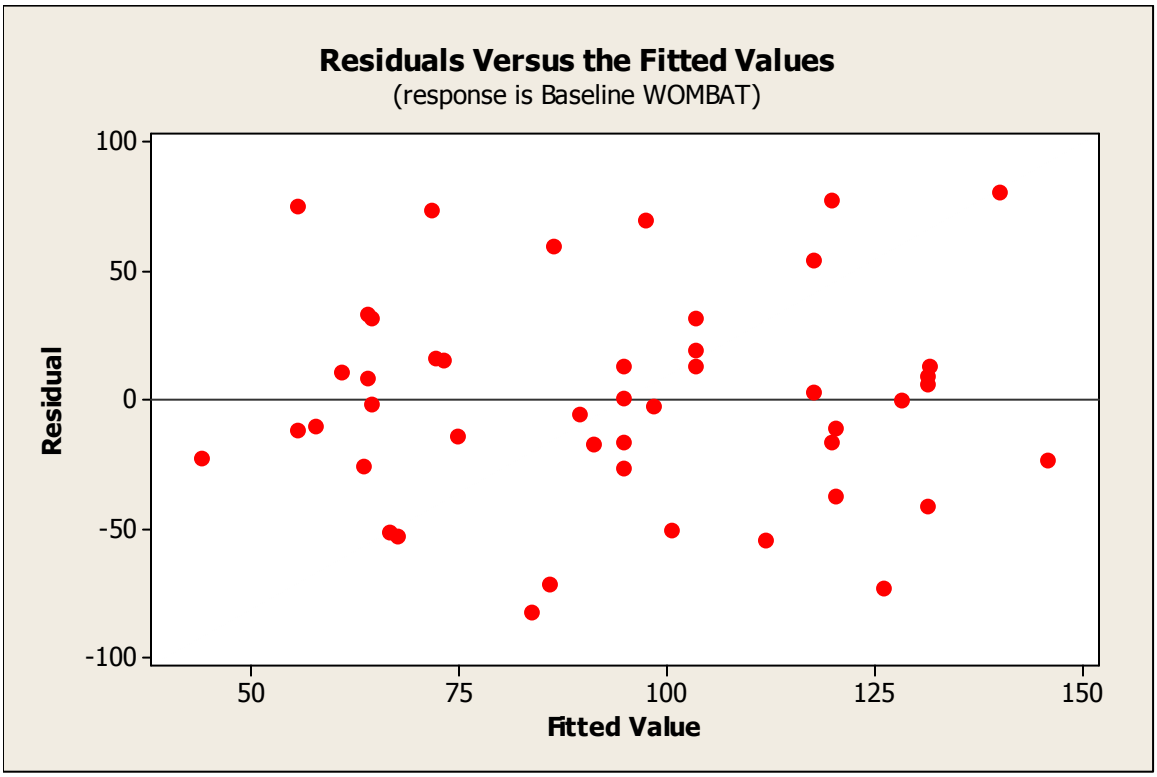


Figure G.2.3.2 Resdiuals vs. Fitted Values for Baseline Regression (no outliers)

G.2.4 Sleep Deprived Regression Plots

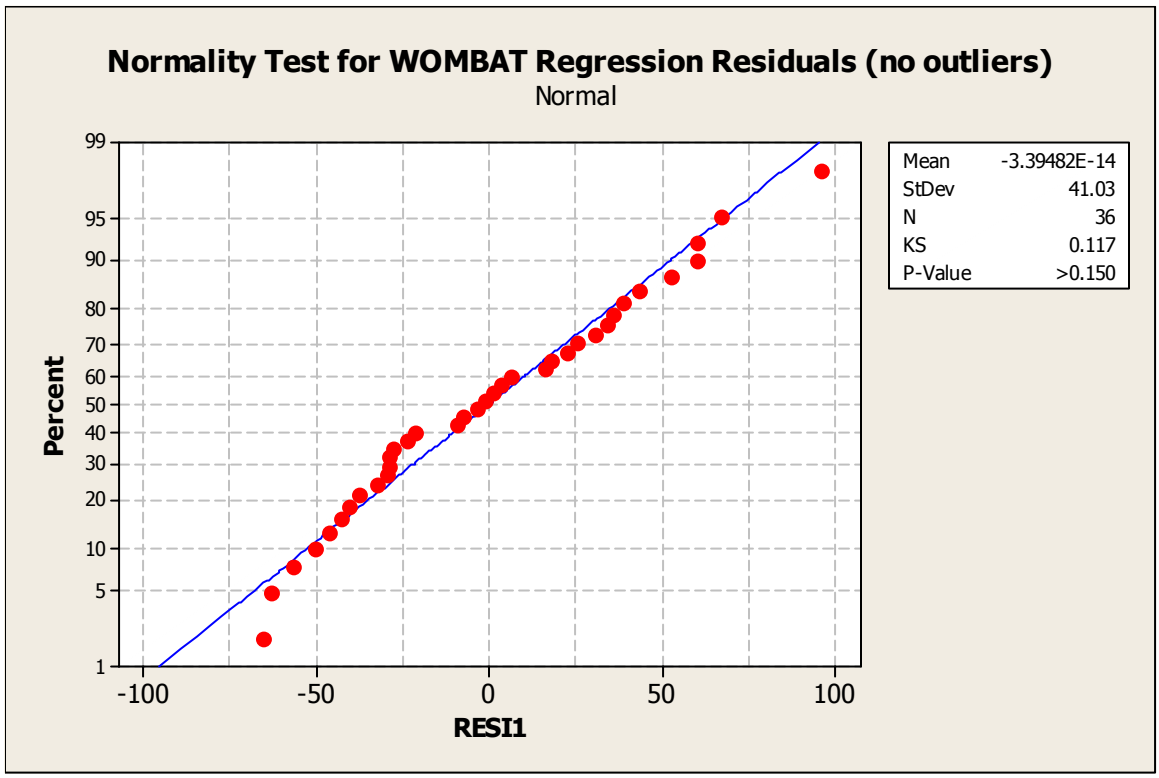


Figure G.2.4.1 Normailty Test for Sleep Deprived Residuals (no outliers)

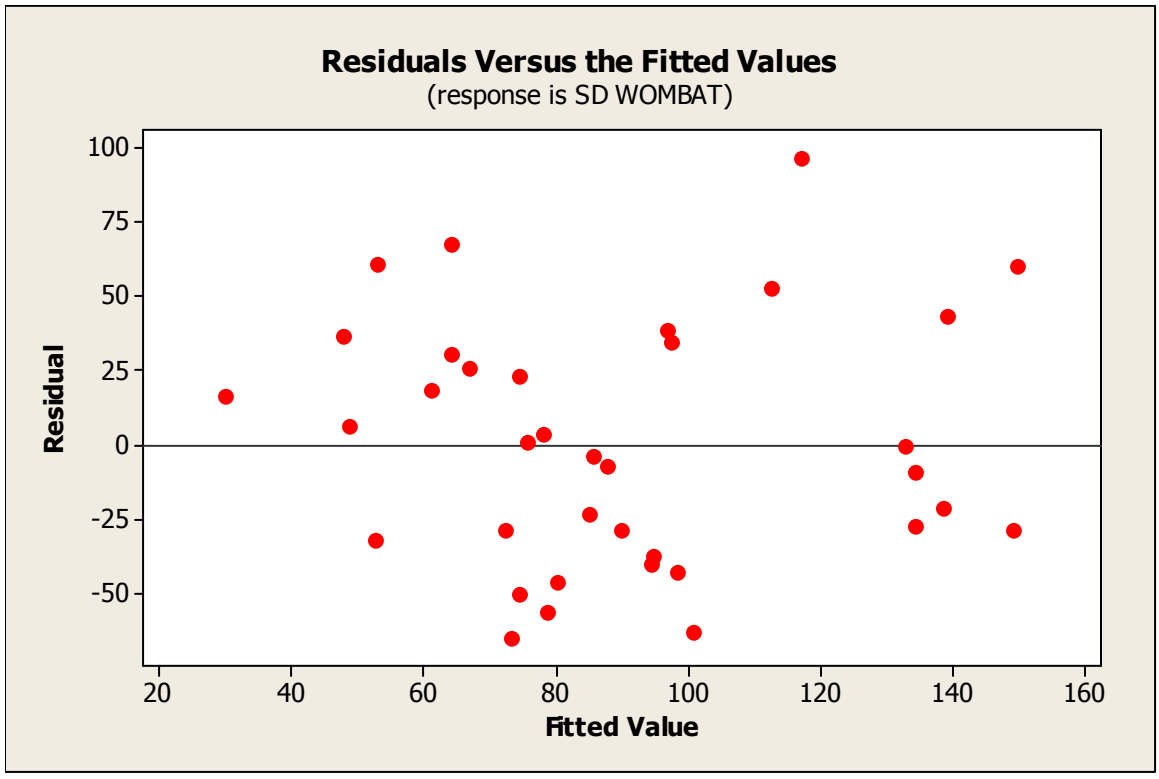


Figure G.2.4.2 Residuals vs. Fitted Values for Sleep Deprived Regression (no outliers)

APPENDIX H. POST HOC SIMPLE LINEAR REGRESSION ANALYSES

(WOMBAT and PVT)

H.1 Baseline WOMBAT Simple Linear Regression Analysis

The regression equation is
Baseline Mean Interval Overall = 6.70 + 0.0356 Time (min)

Predictor	Coef	SE Coef	T	P
Constant	6.7039	0.2346	28.58	0.000
Time (min)	0.035597	0.006374	5.58	0.000

S = 0.381129 R-Sq = 75.7% R-Sq(adj) = 73.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4.5301	4.5301	31.19	0.000
Residual Error	10	1.4526	0.1453		
Total	11	5.9827			

Obs	Time (min)	Baseline Mean Interval Overall	Fit	SE Fit	Residual	St Resid
1	5.0	6.594	6.882	0.207	-0.288	-0.90
2	10.0	7.096	7.060	0.181	0.036	0.11
3	15.0	7.852	7.238	0.157	0.614	1.77
4	20.0	7.460	7.416	0.136	0.045	0.13
5	25.0	7.488	7.594	0.120	-0.106	-0.29
6	30.0	7.919	7.772	0.111	0.147	0.40
7	35.0	7.115	7.950	0.111	-0.835	-2.29R
8	40.0	8.248	8.128	0.120	0.120	0.33
9	45.0	8.223	8.306	0.136	-0.083	-0.23
10	50.0	8.646	8.484	0.157	0.162	0.47
11	55.0	9.067	8.662	0.181	0.405	1.21
12	60.0	8.623	8.840	0.207	-0.217	-0.68

R denotes an observation with a large standardized residual.

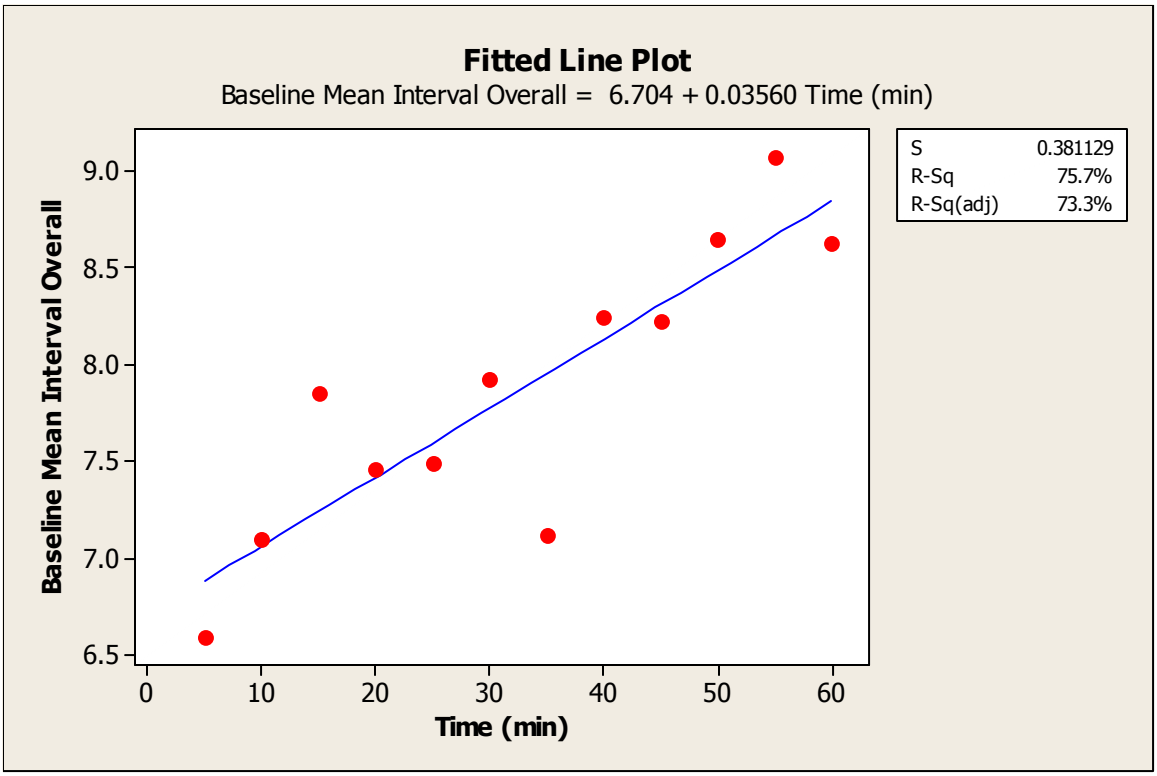


Figure H.1.1 Fitted Line Plot for Baseline WOMBAT over Time

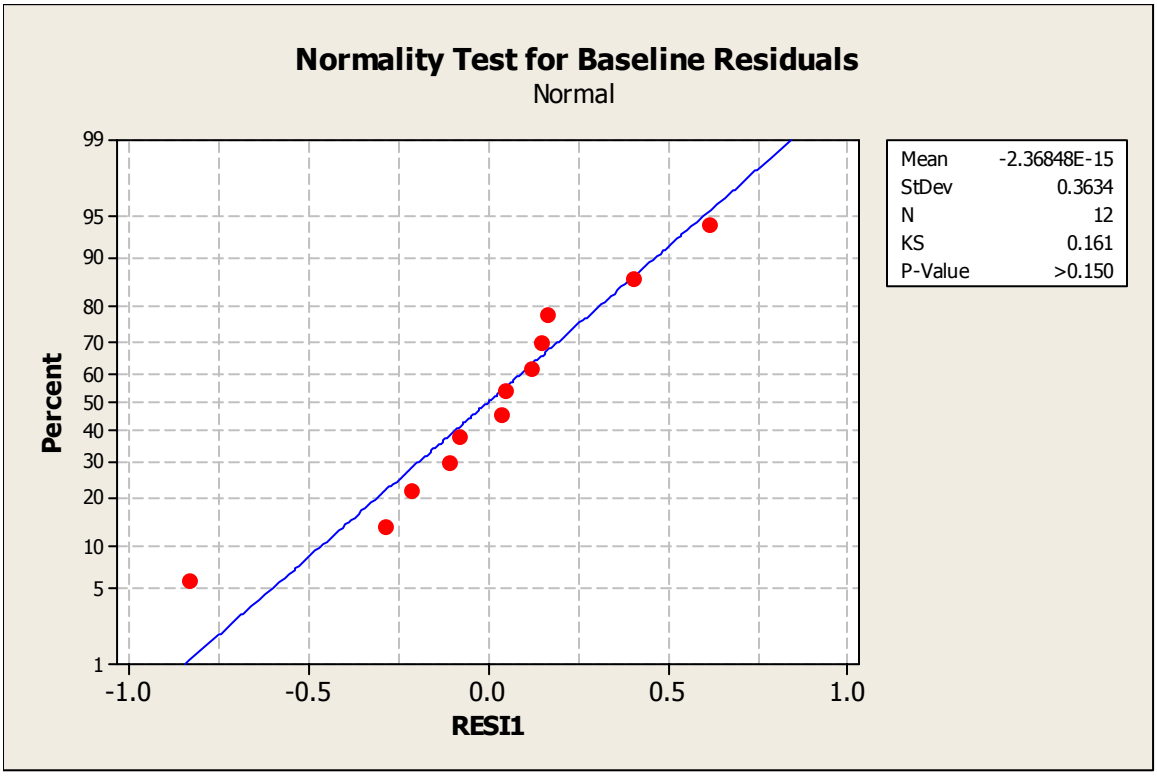


Figure H.1.2 Normality Test for Baseline REgrssion Residuals

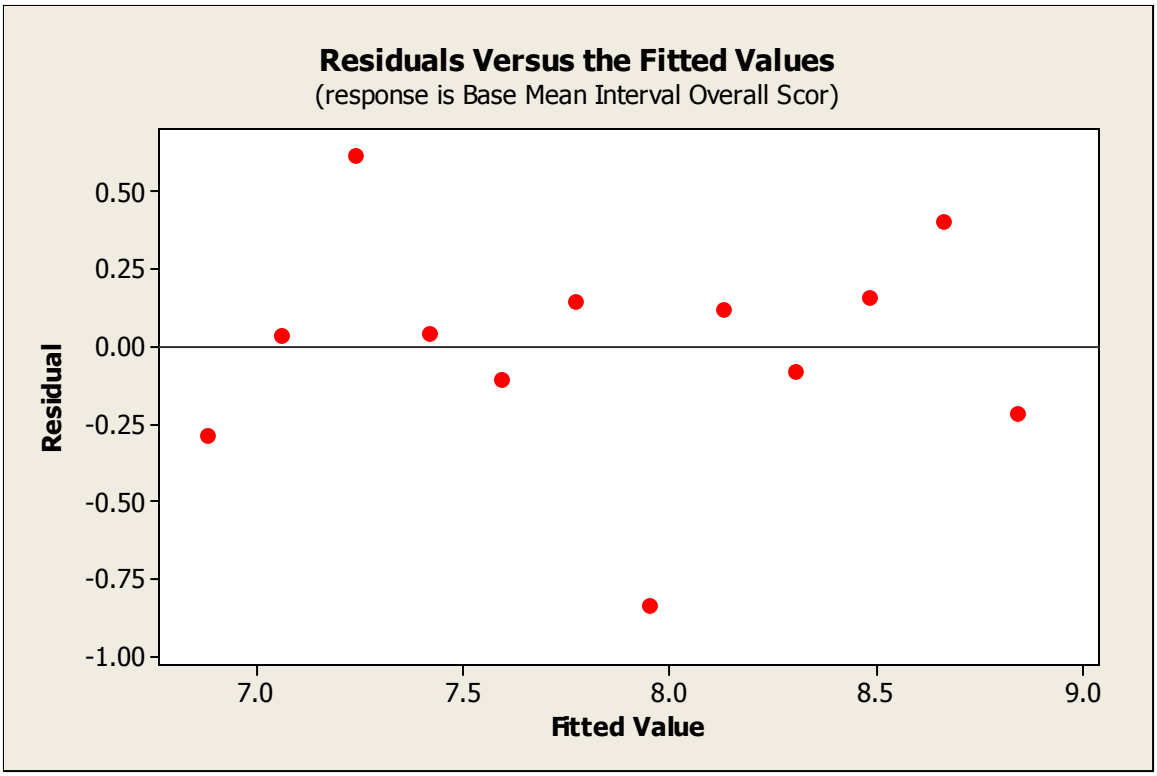


Figure H.1.3 Baseline Residuals vs. Fitted Values

H.2 Sleep DerivedWOMBAT Simple Linear Regression Analysis

```
The regression equation is
SD Mean Interval Overall Scores = 7.43 + 0.0249 Time (min)


Predictor       Coef   SE Coef      T      P
Constant      7.4348    0.3421  21.73  0.000
Time (min)  0.024860  0.009297   2.67  0.023


S = 0.555894   R-Sq = 41.7%   R-Sq(adj) = 35.9%


Analysis of Variance

Source          DF      SS      MS     F      P
Regression       1  2.2094  2.2094  7.15  0.023
Residual Error  10  3.0902  0.3090
Total           11  5.2996


             SD Mean
            Interval
      Time   Overall
Obs   (min)    Scores     Fit  SE Fit  Residual  St Resid
  1    5.0     7.021   7.559   0.302    -0.538     -1.15
  2   10.0     8.052   7.683   0.264     0.369      0.75
  3   15.0     7.879   7.808   0.229     0.072      0.14
  4   20.0     8.346   7.932   0.198     0.414      0.80
  5   25.0     8.775   8.056   0.175     0.719      1.36
```

6	30.0	6.875	8.181	0.162	-1.306	-2.46R
7	35.0	8.460	8.305	0.162	0.156	0.29
8	40.0	8.229	8.429	0.175	-0.200	-0.38
9	45.0	8.654	8.553	0.198	0.101	0.19
10	50.0	9.073	8.678	0.229	0.395	0.78
11	55.0	8.813	8.802	0.264	0.010	0.02
12	60.0	8.735	8.926	0.302	-0.191	-0.41

R denotes an observation with a large standardized residual.

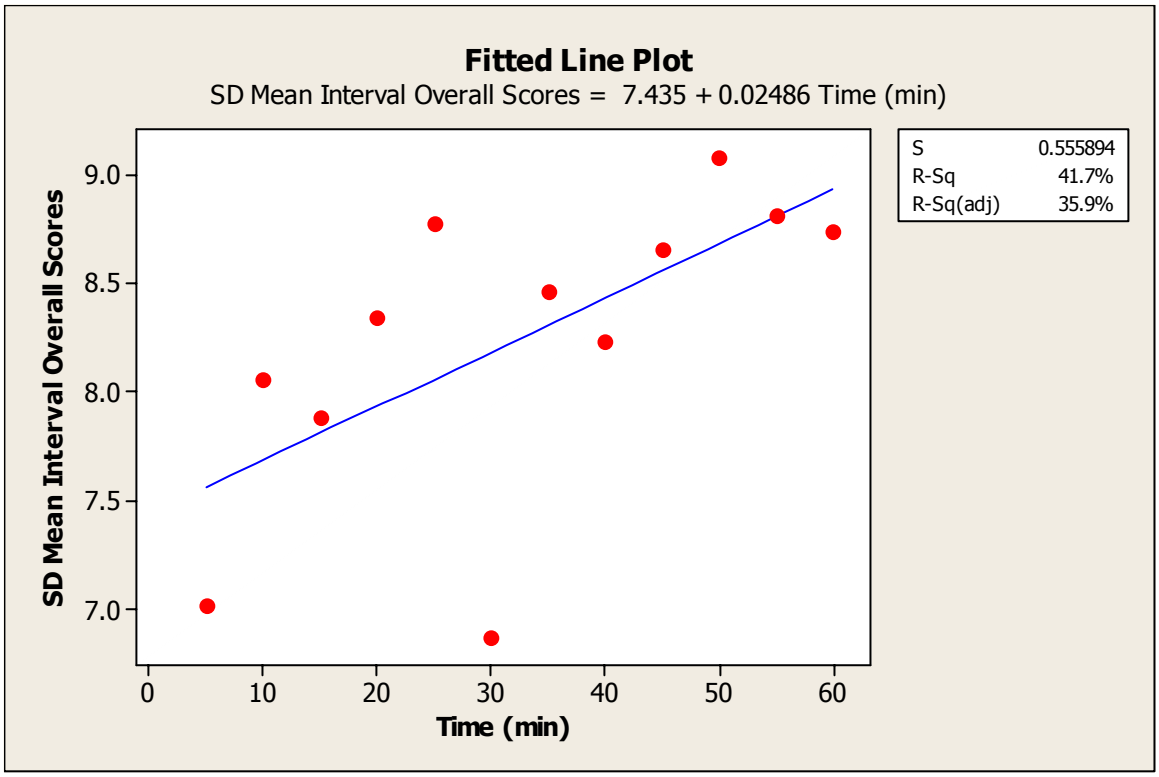


Figure H.2.1 Fitted Line Plot for Sleep Deprived WOMBAT over Time

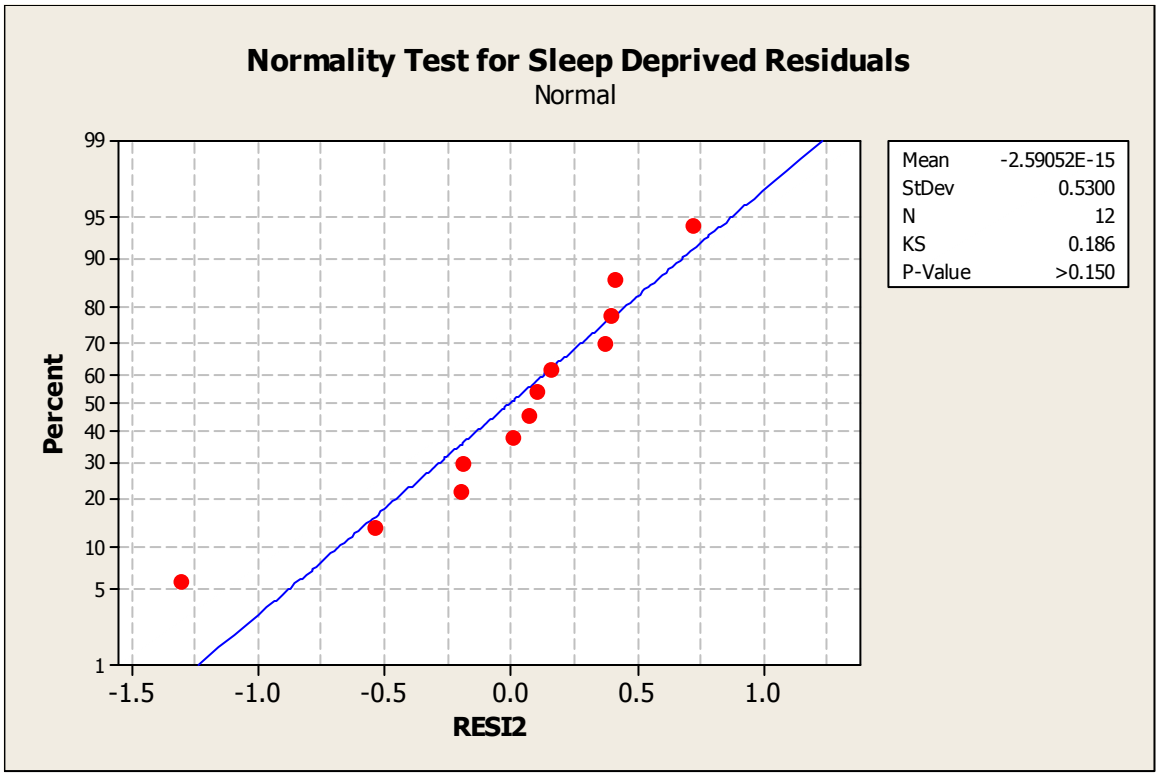


Figure H.2.2 Normatilyt Test for Sleep Deprived Residuals

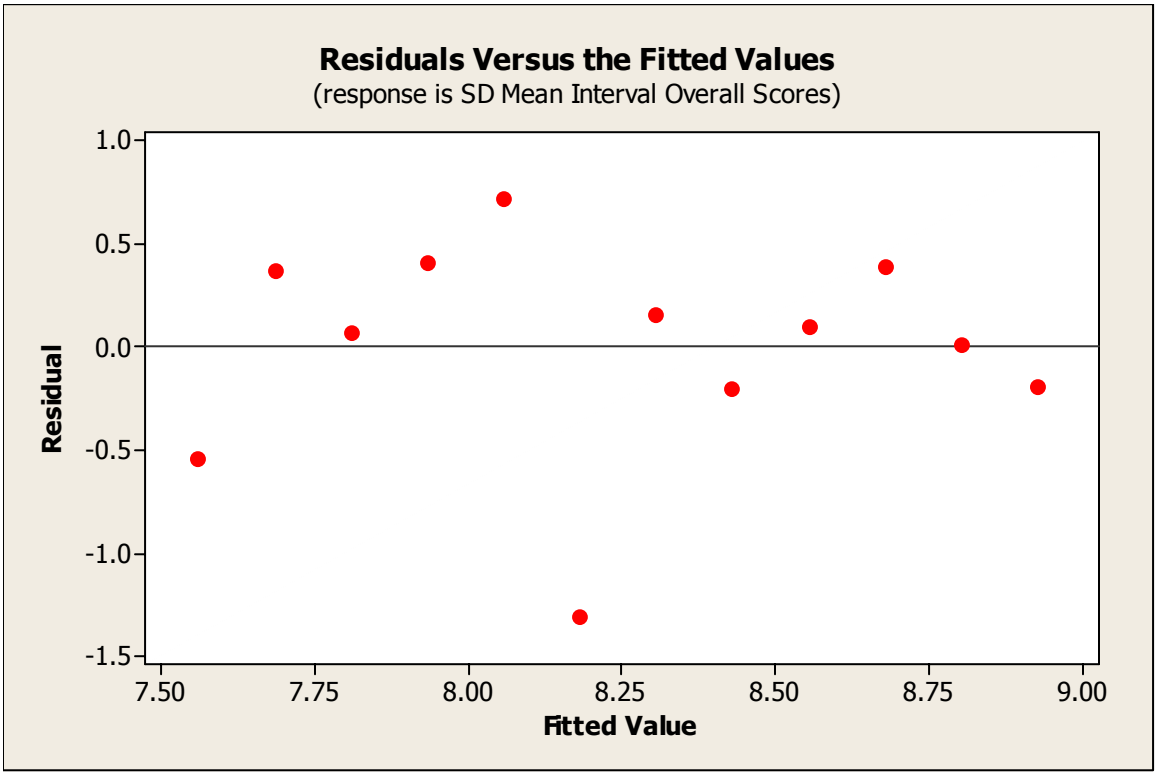


Figure H.2.3 Sleep Deprived Residuals vs. Fitted Values

H.3 Two Factor ANOVA for WOMBAT Score (Interval x Session)

Table H.3.1 Two Factor ANOVA (Interval x Session)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power(a)
session	Sphericity Assumed	42.014	1	42.014	1.308	.259	1.308	.202
	Greenhouse-Geisser	42.014	1.000	42.014	1.308	.259	1.308	.202
	Huynh-Feldt	42.014	1.000	42.014	1.308	.259	1.308	.202
	Lower-bound	42.014	1.000	42.014	1.308	.259	1.308	.202
Error(session)	Sphericity Assumed	1510.208	47	32.132				
	Greenhouse-Geisser	1510.208	47.000	32.132				
	Huynh-Feldt	1510.208	47.000	32.132				
	Lower-bound	1510.208	47.000	32.132				
interval	Sphericity Assumed	418.307	11	38.028	4.491	.000	49.401	1.000
	Greenhouse-Geisser	418.307	5.116	81.770	4.491	.001	22.974	.971
	Huynh-Feldt	418.307	5.815	71.938	4.491	.000	26.114	.983
	Lower-bound	418.307	1.000	418.307	4.491	.039	4.491	.546
Error(interval)	Sphericity Assumed	4377.763	517	8.468				
	Greenhouse-Geisser	4377.763	240.436	18.208				
	Huynh-Feldt	4377.763	273.298	16.018				
	Lower-bound	4377.763	47.000	93.144				
session * interval	Sphericity Assumed	123.246	11	11.204	1.702	.070	18.718	.842
	Greenhouse-Geisser	123.246	8.119	15.179	1.702	.095	13.816	.745
	Huynh-Feldt	123.246	9.972	12.359	1.702	.078	16.969	.812
	Lower-bound	123.246	1.000	123.246	1.702	.198	1.702	.248
Error(session*interval)	Sphericity Assumed	3404.182	517	6.584				
	Greenhouse-Geisser	3404.182	381.614	8.920				
	Huynh-Feldt	3404.182	468.704	7.263				
	Lower-bound	3404.182	47.000	72.429				

Estimated Marginal Means

1. Session

Table H.3.2 Session Estimates

Measure: score

session	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	7.861	.603	6.647	9.075
2	8.243	.668	6.900	9.586

Table H.3.3 Session Pairwise Comparisons

Measure: score

(I) session	(J) session	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
1	2	-.382	.334	.259	-1.054	.290
2	1	.382	.334	.259	-.290	1.054

Based on estimated marginal means
a Adjustment for multiple comparisons: Bonferroni.

Table H.3.4 Session Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power(a)
Pillai's trace	.027	1.308(b)	1.000	47.000	.259	1.308	.202
Wilks' lambda	.973	1.308(b)	1.000	47.000	.259	1.308	.202
Hotelling's trace	.028	1.308(b)	1.000	47.000	.259	1.308	.202
Roy's largest root	.028	1.308(b)	1.000	47.000	.259	1.308	.202

Each F tests the multivariate effect of session. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.
a Computed using alpha = .05
b Exact statistic

2. Interval

Table H.3.5 Internal Estimates

Measure: score

interval	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	6.807	.517	5.768	7.846
2	7.574	.466	6.637	8.511
3	7.866	.528	6.803	8.928
4	7.903	.544	6.809	8.997
5	8.131	.693	6.737	9.526

6	7.397	.692	6.005	8.789
7	7.787	.724	6.330	9.245
8	8.239	.758	6.714	9.763
9	8.439	.771	6.888	9.989
10	8.859	.710	7.431	10.288
11	8.940	.787	7.357	10.522
12	8.679	.813	7.044	10.314

H.3.1 Bonferroni Post Hoc Pairwise Comparisons (WOMBAT by Interval)

Table H.3.1.1 Bonferroni Post Hoc Pairwise Comparisons (WOMBAT by Interval)

Measure: score

(I) interval	(J) interval	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
1	2	-.767	.308	1.000	-1.875	.342
	3	-1.058	.437	1.000	-2.634	.517
	4	-1.096	.391	.482	-2.503	.312
	5	-1.324	.499	.719	-3.123	.475
	6	-.590	.522	1.000	-2.471	1.292
	7	-.980	.485	1.000	-2.728	.768
	8	-1.431	.510	.479	-3.268	.406
	9	-1.631	.514	.176	-3.484	.222
	10	-2.052(*)	.469	.004	-3.740	-.364
	11	-2.132(*)	.566	.030	-4.171	-.093
	12	-1.872	.575	.139	-3.944	.200
2	1	.767	.308	1.000	-.342	1.875
	3	-.292	.352	1.000	-1.560	.976
	4	-.329	.314	1.000	-1.461	.803
	5	-.557	.420	1.000	-2.072	.957
	6	.177	.472	1.000	-1.525	1.879
	7	-.214	.454	1.000	-1.848	1.421
	8	-.665	.477	1.000	-2.385	1.055
	9	-.865	.480	1.000	-2.594	.864
	10	-1.285	.456	.467	-2.930	.359
	11	-1.366	.535	.922	-3.292	.561
	12	-1.105	.589	1.000	-3.228	1.017
3	1	1.058	.437	1.000	-.517	2.634
	2	.292	.352	1.000	-.976	1.560
	4	-.037	.300	1.000	-1.118	1.043
	5	-.266	.386	1.000	-1.656	1.125
	6	.469	.421	1.000	-1.047	1.985
	7	.078	.429	1.000	-1.467	1.624
	8	-.373	.443	1.000	-1.969	1.223
	9	-.573	.470	1.000	-2.268	1.122

	10	-.994	.430	1.000	-2.542	.554
	11	-1.074	.480	1.000	-2.802	.654
	12	-.814	.540	1.000	-2.760	1.133
4	1	1.096	.391	.482	-.312	2.503
	2	.329	.314	1.000	-.803	1.461
	3	.037	.300	1.000	-1.043	1.118
	5	-.228	.354	1.000	-1.502	1.045
	6	.506	.354	1.000	-.768	1.780
	7	.116	.408	1.000	-1.353	1.585
	8	-.335	.428	1.000	-1.877	1.206
	9	-.535	.438	1.000	-2.115	1.044
	10	-.956	.447	1.000	-2.568	.655
	11	-1.036	.507	1.000	-2.862	.789
	12	-.776	.541	1.000	-2.724	1.172
5	1	1.324	.499	.719	-.475	3.123
	2	.557	.420	1.000	-.957	2.072
	3	.266	.386	1.000	-1.125	1.656
	4	.228	.354	1.000	-1.045	1.502
	6	.734	.346	1.000	-.512	1.981
	7	.344	.321	1.000	-.813	1.500
	8	-.107	.332	1.000	-1.305	1.091
	9	-.307	.415	1.000	-1.802	1.187
	10	-.728	.354	1.000	-2.003	.546
	11	-.808	.408	1.000	-2.280	.663
	12	-.548	.450	1.000	-2.170	1.074
6	1	.590	.522	1.000	-1.292	2.471
	2	-.177	.472	1.000	-1.879	1.525
	3	-.469	.421	1.000	-1.985	1.047
	4	-.506	.354	1.000	-1.780	.768
	5	-.734	.346	1.000	-1.981	.512
	7	-.391	.363	1.000	-1.697	.916
	8	-.842	.367	1.000	-2.162	.479
	9	-1.042	.331	.189	-2.234	.151
	10	-1.462(*)	.336	.005	-2.674	-.251
	11	-1.543(*)	.410	.031	-3.019	-.067
	12	-1.282	.444	.384	-2.881	.316
7	1	.980	.485	1.000	-.768	2.728
	2	.214	.454	1.000	-1.421	1.848
	3	-.078	.429	1.000	-1.624	1.467
	4	-.116	.408	1.000	-1.585	1.353
	5	-.344	.321	1.000	-1.500	.813
	6	.391	.363	1.000	-.916	1.697
	8	-.451	.325	1.000	-1.621	.719
	9	-.651	.335	1.000	-1.856	.554
	10	-1.072	.301	.056	-2.155	.011
	11	-1.152	.335	.081	-2.358	.054
	12	-.892	.403	1.000	-2.345	.561
8	1	1.431	.510	.479	-.406	3.268

	2	.665	.477	1.000	-1.055	2.385
	3	.373	.443	1.000	-1.223	1.969
	4	.335	.428	1.000	-1.206	1.877
	5	.107	.332	1.000	-1.091	1.305
	6	.842	.367	1.000	-.479	2.162
	7	.451	.325	1.000	-.719	1.621
	9	-.200	.374	1.000	-1.548	1.148
	10	-.621	.327	1.000	-1.800	.559
	11	-.701	.388	1.000	-2.098	.696
	12	-.441	.394	1.000	-1.861	.980
9	1	1.631	.514	.176	-.222	3.484
	2	.865	.480	1.000	-.864	2.594
	3	.573	.470	1.000	-1.122	2.268
	4	.535	.438	1.000	-1.044	2.115
	5	.307	.415	1.000	-1.187	1.802
	6	1.042	.331	.189	-.151	2.234
	7	.651	.335	1.000	-.554	1.856
	8	.200	.374	1.000	-1.148	1.548
	10	-.421	.317	1.000	-1.563	.721
	11	-.501	.344	1.000	-1.742	.740
	12	-.241	.418	1.000	-1.748	1.267
10	1	2.052(*)	.469	.004	.364	3.740
	2	1.285	.456	.467	-.359	2.930
	3	.994	.430	1.000	-.554	2.542
	4	.956	.447	1.000	-.655	2.568
	5	.728	.354	1.000	-.546	2.003
	6	1.462(*)	.336	.005	.251	2.674
	7	1.072	.301	.056	-.011	2.155
	8	.621	.327	1.000	-.559	1.800
	9	.421	.317	1.000	-.721	1.563
	11	-.080	.277	1.000	-1.077	.916
	12	.180	.331	1.000	-1.012	1.372
11	1	2.132(*)	.566	.030	.093	4.171
	2	1.366	.535	.922	-.561	3.292
	3	1.074	.480	1.000	-.654	2.802
	4	1.036	.507	1.000	-.789	2.862
	5	.808	.408	1.000	-.663	2.280
	6	1.543(*)	.410	.031	.067	3.019
	7	1.152	.335	.081	-.054	2.358
	8	.701	.388	1.000	-.696	2.098
	9	.501	.344	1.000	-.740	1.742
	10	.080	.277	1.000	-.916	1.077
	12	.260	.300	1.000	-.820	1.341
12	1	1.872	.575	.139	-.200	3.944
	2	1.105	.589	1.000	-1.017	3.228
	3	.814	.540	1.000	-1.133	2.760
	4	.776	.541	1.000	-1.172	2.724
	5	.548	.450	1.000	-1.074	2.170

6	1.282	.444	.384	-.316	2.881
7	.892	.403	1.000	-.561	2.345
8	.441	.394	1.000	-.980	1.861
9	.241	.418	1.000	-1.267	1.748
10	-.180	.331	1.000	-1.372	1.012
11	-.260	.300	1.000	-1.341	.820

Based on estimated marginal means
* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

Table H.3.1.2 Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power(a)
Pillai's trace	.572	4.490(b)	11.000	37.000	.000	49.388	.997
Wilks' lambda	.428	4.490(b)	11.000	37.000	.000	49.388	.997
Hotelling's trace	1.335	4.490(b)	11.000	37.000	.000	49.388	.997
Roy's largest root	1.335	4.490(b)	11.000	37.000	.000	49.388	.997

Each F tests the multivariate effect of interval. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.
a Computed using alpha = .05
b Exact statistic

H.3.2 Tukey HSD Post Hoc Pairwise Comparisons (WOMBAT by Interval)

Table H.3.2.1 Tukey HSD Post Hoc Pairwise Comparisons (WOMBAT by Interval)

Session	Score	vs.	Session	Score	Difference	**p<.05**
INT 1	6.807		INT 2	7.574	-0.76700000000000	ns
INT 1	6.807		INT 3	7.866	-1.05900000000000	ns
INT 1	6.807		INT 4	7.903	-1.09600000000000	ns
INT 1	6.807		INT 5	8.131	-1.32400000000000	ns
INT 1	6.807		INT 6	7.397	-0.59000000000000	ns
INT 1	6.807		INT 7	7.787	-0.98000000000000	ns
INT 1	6.807		INT 8	8.239	-1.43200000000000	ns
INT 1	6.807		INT 9	8.439	-1.63200000000000	ns
INT 1	6.807		INT 10	8.859	-2.05200000000000	sig
INT 1	6.807		INT 11	8.94	-2.13300000000000	sig
INT 1	6.807		INT 12	8.679	-1.87200000000000	ns
INT 2	7.574		INT 3	7.866	-0.29200000000000	ns
INT 2	7.574		INT 4	7.903	-0.32900000000000	ns
INT 2	7.574		INT 5	8.131	-0.55700000000000	ns
INT 2	7.574		INT 6	7.397	0.17700000000000	ns
INT 2	7.574		INT 7	7.787	-0.21300000000000	ns
INT 2	7.574		INT 8	8.239	-0.66500000000000	ns
INT 2	7.574		INT 9	8.439	-0.86500000000000	ns
INT 2	7.574		INT 10	8.859	-1.28500000000000	ns
INT 2	7.574		INT 11	8.94	-1.36600000000000	ns
INT 2	7.574		INT 12	8.679	-1.10500000000000	ns

INT 3	7.866	INT 4	7.903	-0.03700000000000	ns
INT 3	7.866	INT 5	8.131	-0.26500000000000	ns
INT 3	7.866	INT 6	7.397	0.46900000000000	ns
INT 3	7.866	INT 7	7.787	0.07900000000000	ns
INT 3	7.866	INT 8	8.239	-0.37300000000000	ns
INT 3	7.866	INT 9	8.439	-0.57300000000000	ns
INT 3	7.866	INT 10	8.859	-0.99300000000000	ns
INT 3	7.866	INT 11	8.94	-1.07400000000000	ns
INT 3	7.866	INT 12	8.679	-0.81300000000000	ns
INT 4	7.903	INT 5	8.131	-0.22800000000000	ns
INT 4	7.903	INT 6	7.397	0.50600000000000	ns
INT 4	7.903	INT 7	7.787	0.11600000000000	ns
INT 4	7.903	INT 8	8.239	-0.33600000000000	ns
INT 4	7.903	INT 9	8.439	-0.53600000000000	ns
INT 4	7.903	INT 10	8.859	-0.95600000000000	ns
INT 4	7.903	INT 11	8.94	-1.03700000000000	ns
INT 4	7.903	INT 12	8.679	-0.77600000000000	ns
INT 5	8.131	INT 6	7.397	0.73400000000000	ns
INT 5	8.131	INT 7	7.787	0.34400000000000	ns
INT 5	8.131	INT 8	8.239	-0.10800000000000	ns
INT 5	8.131	INT 9	8.439	-0.30800000000000	ns
INT 5	8.131	INT 10	8.859	-0.72800000000000	ns
INT 5	8.131	INT 11	8.94	-0.80900000000000	ns
INT 5	8.131	INT 12	8.679	-0.54800000000000	ns
INT 6	7.397	INT 7	7.787	-0.39000000000000	ns
INT 6	7.397	INT 8	8.239	-0.84200000000000	ns
INT 6	7.397	INT 9	8.439	-1.04200000000000	ns
INT 6	7.397	INT 10	8.859	-1.46200000000000	ns
INT 6	7.397	INT 11	8.94	-1.54300000000000	ns
INT 6	7.397	INT 12	8.679	-1.28200000000000	ns
INT 7	7.787	INT 8	8.239	-0.45200000000000	ns
INT 7	7.787	INT 9	8.439	-0.65200000000000	ns
INT 7	7.787	INT 10	8.859	-1.07200000000000	ns
INT 7	7.787	INT 11	8.94	-1.15300000000000	ns
INT 7	7.787	INT 12	8.679	-0.89200000000000	ns
INT 8	8.239	INT 9	8.439	-0.20000000000000	ns
INT 8	8.239	INT 10	8.859	-0.62000000000000	ns
INT 8	8.239	INT 11	8.94	-0.70100000000000	ns
INT 8	8.239	INT 12	8.679	-0.44000000000000	ns
INT 9	8.439	INT 10	8.859	-0.42000000000000	ns
INT 9	8.439	INT 11	8.94	-0.50100000000000	ns
INT 9	8.439	INT 12	8.679	-0.24000000000000	ns
INT 10	8.859	INT 11	8.94	-0.08100000000000	ns
INT 10	8.859	INT 12	8.679	0.18000000000000	ns
INT 11	8.940	INT 12	8.679	0.26100000000000	ns

H.4 PVT Simple Linear Regression Analysis

H.4.1 Linear Regression Analysis: mean RT versus session

```
The regression equation is
mean RT = 222.0 + 5.284 session


S = 20.4397   R-Sq = 59.0%   R-Sq(adj) = 55.9%


Analysis of Variance

Source       DF       SS       MS      F      P
Regression    1   7817.9  7817.88  18.71  0.001
Error        13   5431.1   417.78
Total        14  13249.0
```

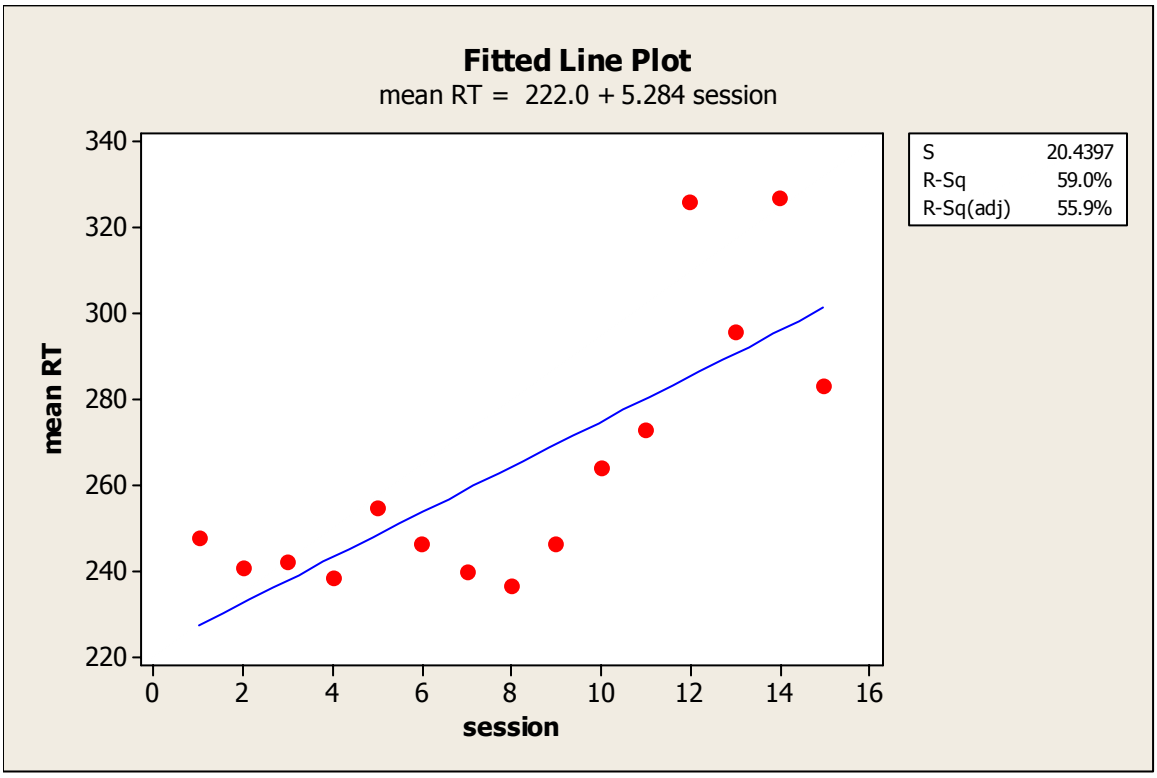


Figure H.4.1.1 Linear Regression Analysis: mean RT versus session Fitted Line Plot

H.4.2 Quadratic Regression Analysis: mean RT versus session

```
The regression equation is
mean RT = 246.7 - 3.437 session + 0.5451 session**2


S = 18.7205   R-Sq = 68.3%   R-Sq(adj) = 63.0%


Analysis of Variance
```

```
Source       DF        SS        MS       F      P
Regression    2    9043.6   4521.78   12.90  0.001
Error        12    4205.5    350.46
Total        14   13249.0
```

```
Sequential Analysis of Variance

Source      DF       SS      F      P
Linear       1  7817.88  18.71  0.001
Quadratic    1  1225.68   3.50  0.086
```

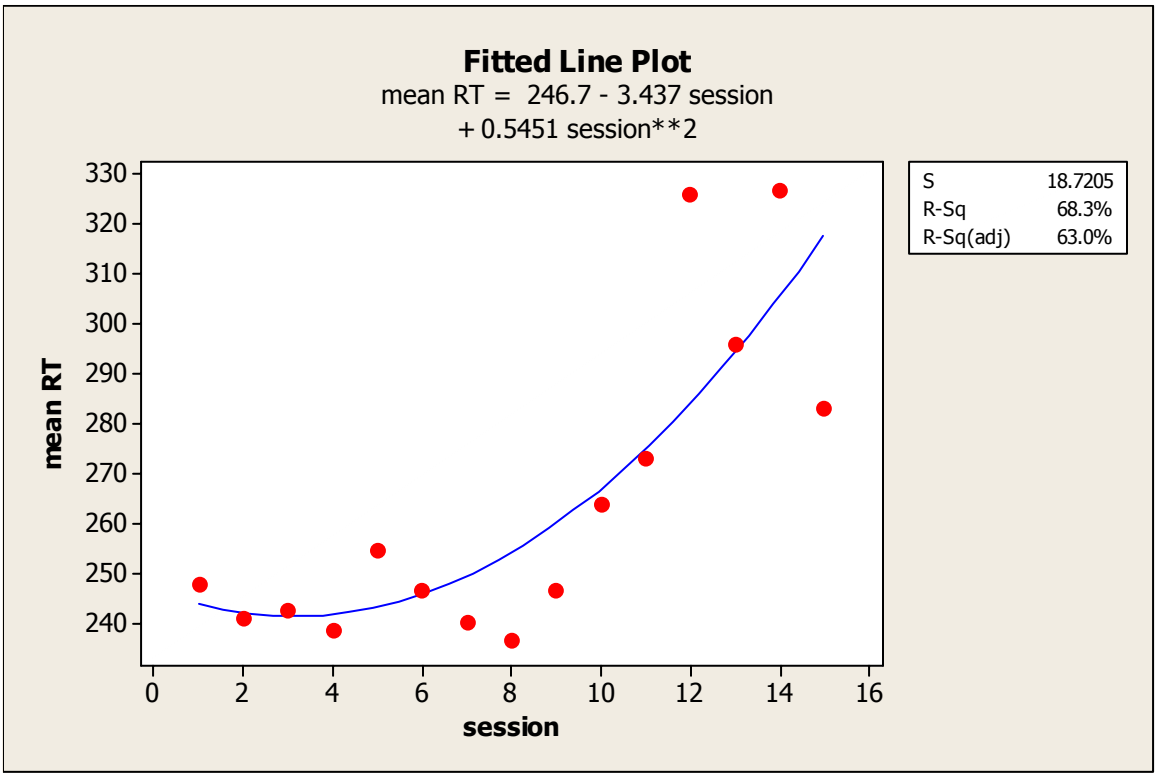


Figure H.4.2.1 Quadratic Regression Analysis: mean RT versus session Fitted Line Plot

H.5 ANOVA for Mean RT by Session

Table H.5.1 ANOVA for Mean RT by Session Within-Subjects Factors

Measure: meanrt

session	Dependent Variable
1	RT1
2	RT2
3	RT3
4	RT4
5	RT5
6	RT6
7	RT7
8	RT8
9	RT9
10	RT10
11	RT11
12	RT12
13	RT13
14	RT14
15	RT15

Table H.5.2 ANOVA for Mean RT by Session Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
session	Pillai's Trace	.702	5.380(a)	14.000	32.000	.000
	Wilks' Lambda	.298	5.380(a)	14.000	32.000	.000
	Hotelling's Trace	2.354	5.380(a)	14.000	32.000	.000
	Roy's Largest Root	2.354	5.380(a)	14.000	32.000	.000

a Exact statistic
b Design: Intercept Within Subjects Design: session

Table H.5.3 ANOVA for Mean RT by Session Mauchly's Test of Sphericity(b)

Measure: meanrt

					Epsilon(a)		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
session	.000	1321.751	104	.000	.141	.147	.071

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.
a May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
b Design: Intercept Within Subjects Design: session

Table H.5.4 ANOVA for Mean RT by Session Tests of Within-Subjects Effects

Measure: meanrt

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
session	Sphericity Assumed	1245386.083	14	88956.149	6.614	.000
	Greenhouse-Geisser	1245386.083	1.970	632245.046	6.614	.002
	Huynh-Feldt	1245386.083	2.059	604775.958	6.614	.002
	Lower-bound	1245386.083	1.000	1245386.083	6.614	.013
Error(session)	Sphericity Assumed	8473723.438	630	13450.355		
	Greenhouse-Geisser	8473723.438	88.640	95596.766		
	Huynh-Feldt	8473723.438	92.666	91443.382		
	Lower-bound	8473723.438	45.000	188304.965		

Table H.5.5 ANOVA for Mean RT by Session Tests of Within-Subjects Contrasts

Measure: meanrt

Source	session	Type III Sum of Squares	df	Mean Square	F	Sig.
session	Linear	694364.967	1	694364.967	20.734	.000
	Quadratic	88855.311	1	88855.311	17.633	.000
	Cubic	82196.664	1	82196.664	3.806	.057
	Order 4	207744.358	1	207744.358	8.383	.006
	Order 5	68908.740	1	68908.740	11.372	.002
	Order 6	6727.772	1	6727.772	.650	.424
	Order 7	18291.596	1	18291.596	.843	.363
	Order 8	2152.023	1	2152.023	.153	.697
	Order 9	20.011	1	20.011	.003	.958
	Order 10	1804.512	1	1804.512	.169	.683
	Order 11	33387.103	1	33387.103	2.384	.130
	Order 12	17044.120	1	17044.120	1.485	.229
	Order 13	23800.742	1	23800.742	3.459	.069
	Order 14	88.164	1	88.164	.083	.775
Error(session)	Linear	1506987.058	45	33488.601		
	Quadratic	226767.605	45	5039.280		
	Cubic	971840.242	45	21596.450		
	Order 4	1115236.930	45	24783.043		
	Order 5	272669.446	45	6059.321		
	Order 6	465458.610	45	10343.525		
	Order 7	976583.824	45	21701.863		
	Order 8	632265.740	45	14050.350		
	Order 9	321179.496	45	7137.322		
	Order 10	480431.988	45	10676.266		
	Order 11	630209.238	45	14004.650		

	Order 12	516397.105	45	11475.491		
	Order 13	309653.776	45	6881.195		
	Order 14	48042.380	45	1067.608		

Table H.5.6 ANOVA for Mean RT by Session Tests of Between-Subjects Effects

Measure: meanrt
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	53487739.515	1	53487739.515	859.642	.000
Error	2799944.004	45	62220.978		

Estimated Marginal Means

Table H.5.7 Session Estimated Marginal Means

Measure: meanrt

session	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	252.946	6.130	240.599	265.293
2	247.010	5.317	236.302	257.718
3	250.568	6.677	237.120	264.016
4	243.854	5.238	233.305	254.403
5	261.249	6.955	247.240	275.258
6	249.903	4.900	240.033	259.773
7	249.201	7.500	234.094	264.307
8	239.557	4.751	229.987	249.126
9	252.238	5.985	240.184	264.292
10	278.958	12.548	253.685	304.230
11	281.610	10.195	261.077	302.143
12	372.551	53.412	264.973	480.128
13	354.260	36.813	280.115	428.405
14	343.851	20.086	303.397	384.306
15	298.566	15.762	266.819	330.312

H.5.1 Bonferroni Post Hoc Pairwise Comparisons (Mean RT by Session)

Table H.5.1.1 Bonferroni Post Hoc Pairwise Comparisons (Mean RT by Session)

Measure: meanrt

(I) session	(J) session	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound

1	2	5.936	3.982	1.000	-9.068	20.939
	3	2.378	4.719	1.000	-15.402	20.158
	4	9.092	4.085	1.000	-6.302	24.485
	5	-8.303	4.026	1.000	-23.472	6.866
	6	3.043	4.641	1.000	-14.445	20.530
	7	3.745	6.086	1.000	-19.185	26.676
	8	13.389	4.038	.190	-1.825	28.604
	9	.708	4.420	1.000	-15.948	17.364
	10	-26.012	11.463	1.000	-69.205	17.181
	11	-28.664	9.402	.403	-64.090	6.761
	12	-119.605	52.357	1.000	-316.886	77.676
	13	-101.314	35.579	.695	-235.374	32.746
	14	-90.905(*)	18.537	.001	-160.753	-21.057
	15	-45.620	14.981	.407	-102.067	10.828
2	1	-5.936	3.982	1.000	-20.939	9.068
	3	-3.558	5.517	1.000	-24.347	17.232
	4	3.156	3.652	1.000	-10.606	16.918
	5	-14.238	4.478	.280	-31.112	2.635
	6	-2.893	3.188	1.000	-14.904	9.119
	7	-2.190	6.399	1.000	-26.301	21.920
	8	7.454	3.195	1.000	-4.586	19.493
	9	-5.228	4.355	1.000	-21.638	11.183
	10	-31.948	10.780	.509	-72.567	8.672
	11	-34.600(*)	7.486	.003	-62.805	-6.395
	12	-125.540	51.879	1.000	-321.021	69.940
	13	-107.250	34.977	.384	-239.043	24.543
	14	-96.841(*)	18.056	.000	-164.874	-28.807
	15	-51.555	14.174	.074	-104.964	1.853
3	1	-2.378	4.719	1.000	-20.158	15.402
	2	3.558	5.517	1.000	-17.232	24.347
	4	6.713	4.820	1.000	-11.448	24.875
	5	-10.681	3.988	1.000	-25.708	4.347
	6	.665	6.015	1.000	-22.001	23.330
	7	1.367	7.466	1.000	-26.765	29.499
	8	11.011	5.499	1.000	-9.707	31.730
	9	-1.670	5.975	1.000	-24.183	20.843
	10	-28.390	12.345	1.000	-74.905	18.125
	11	-31.042	10.258	.429	-69.695	7.610
	12	-121.983	52.938	1.000	-321.453	77.488
	13	-103.692	35.224	.537	-236.415	29.031
	14	-93.283(*)	19.069	.001	-165.136	-21.431
	15	-47.998	15.334	.322	-105.777	9.782
4	1	-9.092	4.085	1.000	-24.485	6.302
	2	-3.156	3.652	1.000	-16.918	10.606
	3	-6.713	4.820	1.000	-24.875	11.448
	5	-17.394(*)	3.705	.003	-31.354	-3.435
	6	-6.049	3.215	1.000	-18.162	6.065
	7	-5.346	5.909	1.000	-27.611	16.919

	8	4.298	2.886	1.000	-6.575	15.171
	9	-8.383	4.224	1.000	-24.301	7.534
	10	-35.103	11.811	.497	-79.609	9.402
	11	-37.756(*)	8.902	.011	-71.299	-4.213
	12	-128.696	52.798	1.000	-327.637	70.244
	13	-110.406	34.864	.291	-241.771	20.960
	14	-99.997(*)	18.165	.000	-168.444	-31.550
	15	-54.711(*)	13.483	.020	-105.516	-3.907
5	1	8.303	4.026	1.000	-6.866	23.472
	2	14.238	4.478	.280	-2.635	31.112
	3	10.681	3.988	1.000	-4.347	25.708
	4	17.394(*)	3.705	.003	3.435	31.354
	6	11.346	4.586	1.000	-5.935	28.626
	7	12.048	5.563	1.000	-8.912	33.008
	8	21.692(*)	4.097	.000	6.256	37.129
	9	9.011	4.530	1.000	-8.057	26.078
	10	-17.709	11.355	1.000	-60.495	25.077
	11	-20.362	8.911	1.000	-53.939	13.216
	12	-111.302	52.170	1.000	-307.876	85.272
	13	-93.012	33.812	.896	-220.416	34.393
	14	-82.602(*)	17.528	.002	-148.649	-16.556
	15	-37.317	13.147	.712	-86.855	12.221
6	1	-3.043	4.641	1.000	-20.530	14.445
	2	2.893	3.188	1.000	-9.119	14.904
	3	-.665	6.015	1.000	-23.330	22.001
	4	6.049	3.215	1.000	-6.065	18.162
	5	-11.346	4.586	1.000	-28.626	5.935
	7	.702	6.123	1.000	-22.369	23.774
	8	10.347	2.970	.117	-.845	21.538
	9	-2.335	3.717	1.000	-16.339	11.670
	10	-29.055	11.489	1.000	-72.345	14.236
	11	-31.707(*)	7.783	.019	-61.033	-2.382
	12	-122.648	52.752	1.000	-321.416	76.121
	13	-104.357	34.857	.469	-235.696	26.982
	14	-93.948(*)	18.310	.001	-162.939	-24.958
	15	-48.663	13.141	.061	-98.178	.852
7	1	-3.745	6.086	1.000	-26.676	19.185
	2	2.190	6.399	1.000	-21.920	26.301
	3	-1.367	7.466	1.000	-29.499	26.765
	4	5.346	5.909	1.000	-16.919	27.611
	5	-12.048	5.563	1.000	-33.008	8.912
	6	-.702	6.123	1.000	-23.774	22.369
	8	9.644	5.769	1.000	-12.094	31.383
	9	-3.037	5.500	1.000	-23.759	17.685
	10	-29.757	8.985	.193	-63.614	4.099
	11	-32.410	10.085	.255	-70.410	5.591
	12	-123.350	52.542	1.000	-321.326	74.626
	13	-105.060	35.280	.490	-237.993	27.874

	14	-94.650(*)	18.821	.001	-165.566	-23.734
	15	-49.365	13.553	.073	-100.434	1.704
8	1	-13.389	4.038	.190	-28.604	1.825
	2	-7.454	3.195	1.000	-19.493	4.586
	3	-11.011	5.499	1.000	-31.730	9.707
	4	-4.298	2.886	1.000	-15.171	6.575
	5	-21.692(*)	4.097	.000	-37.129	-6.256
	6	-10.347	2.970	.117	-21.538	.845
	7	-9.644	5.769	1.000	-31.383	12.094
	9	-12.681	3.702	.139	-26.632	1.269
	10	-39.401	11.448	.132	-82.537	3.734
	11	-42.054(*)	8.451	.001	-73.896	-10.212
	12	-132.994	52.627	1.000	-331.291	65.303
	13	-114.704	35.397	.236	-248.079	18.672
	14	-104.295(*)	18.016	.000	-172.180	-36.410
	15	-59.009(*)	14.035	.013	-111.893	-6.125
9	1	-.708	4.420	1.000	-17.364	15.948
	2	5.228	4.355	1.000	-11.183	21.638
	3	1.670	5.975	1.000	-20.843	24.183
	4	8.383	4.224	1.000	-7.534	24.301
	5	-9.011	4.530	1.000	-26.078	8.057
	6	2.335	3.717	1.000	-11.670	16.339
	7	3.037	5.500	1.000	-17.685	23.759
	8	12.681	3.702	.139	-1.269	26.632
	10	-26.720	10.177	1.000	-65.066	11.625
	11	-29.372	7.916	.060	-59.200	.455
	12	-120.313	52.264	1.000	-317.242	76.616
	13	-102.022	35.351	.627	-235.223	31.178
	14	-91.613(*)	17.980	.001	-159.360	-23.866
	15	-46.328	14.041	.199	-99.234	6.578
10	1	26.012	11.463	1.000	-17.181	69.205
	2	31.948	10.780	.509	-8.672	72.567
	3	28.390	12.345	1.000	-18.125	74.905
	4	35.103	11.811	.497	-9.402	79.609
	5	17.709	11.355	1.000	-25.077	60.495
	6	29.055	11.489	1.000	-14.236	72.345
	7	29.757	8.985	.193	-4.099	63.614
	8	39.401	11.448	.132	-3.734	82.537
	9	26.720	10.177	1.000	-11.625	65.066
	11	-2.652	11.570	1.000	-46.248	40.943
	12	-93.593	46.672	1.000	-269.452	82.266
	13	-75.302	34.525	1.000	-205.390	54.785
	14	-64.893	20.444	.285	-141.925	12.139
	15	-19.608	16.971	1.000	-83.554	44.338
11	1	28.664	9.402	.403	-6.761	64.090
	2	34.600(*)	7.486	.003	6.395	62.805
	3	31.042	10.258	.429	-7.610	69.695
	4	37.756(*)	8.902	.011	4.213	71.299

	5	20.362	8.911	1.000	-13.216	53.939
	6	31.707(*)	7.783	.019	2.382	61.033
	7	32.410	10.085	.255	-5.591	70.410
	8	42.054(*)	8.451	.001	10.212	73.896
	9	29.372	7.916	.060	-.455	59.200
	10	2.652	11.570	1.000	-40.943	46.248
	12	-90.940	49.644	1.000	-277.997	96.116
	13	-72.650	33.445	1.000	-198.670	53.370
	14	-62.241	19.092	.223	-134.179	9.697
	15	-16.955	14.810	1.000	-72.759	38.848
12	1	119.605	52.357	1.000	-77.676	316.886
	2	125.540	51.879	1.000	-69.940	321.021
	3	121.983	52.938	1.000	-77.488	321.453
	4	128.696	52.798	1.000	-70.244	327.637
	5	111.302	52.170	1.000	-85.272	307.876
	6	122.648	52.752	1.000	-76.121	321.416
	7	123.350	52.542	1.000	-74.626	321.326
	8	132.994	52.627	1.000	-65.303	331.291
	9	120.313	52.264	1.000	-76.616	317.242
	10	93.593	46.672	1.000	-82.266	269.452
	11	90.940	49.644	1.000	-96.116	277.997
	13	18.290	41.617	1.000	-138.520	175.101
	14	28.700	52.433	1.000	-168.866	226.265
	15	73.985	51.855	1.000	-121.404	269.374
13	1	101.314	35.579	.695	-32.746	235.374
	2	107.250	34.977	.384	-24.543	239.043
	3	103.692	35.224	.537	-29.031	236.415
	4	110.406	34.864	.291	-20.960	241.771
	5	93.012	33.812	.896	-34.393	220.416
	6	104.357	34.857	.469	-26.982	235.696
	7	105.060	35.280	.490	-27.874	237.993
	8	114.704	35.397	.236	-18.672	248.079
	9	102.022	35.351	.627	-31.178	235.223
	10	75.302	34.525	1.000	-54.785	205.390
	11	72.650	33.445	1.000	-53.370	198.670
	12	-18.290	41.617	1.000	-175.101	138.520
	14	10.409	36.430	1.000	-126.860	147.678
	15	55.695	29.995	1.000	-57.327	168.716
14	1	90.905(*)	18.537	.001	21.057	160.753
	2	96.841(*)	18.056	.000	28.807	164.874
	3	93.283(*)	19.069	.001	21.431	165.136
	4	99.997(*)	18.165	.000	31.550	168.444
	5	82.602(*)	17.528	.002	16.556	148.649
	6	93.948(*)	18.310	.001	24.958	162.939
	7	94.650(*)	18.821	.001	23.734	165.566
	8	104.295(*)	18.016	.000	36.410	172.180
	9	91.613(*)	17.980	.001	23.866	159.360
	10	64.893	20.444	.285	-12.139	141.925

	11	62.241	19.092	.223	-9.697	134.179
	12	-28.700	52.433	1.000	-226.265	168.866
	13	-10.409	36.430	1.000	-147.678	126.860
	15	45.285	19.431	1.000	-27.931	118.502
15	1	45.620	14.981	.407	-10.828	102.067
	2	51.555	14.174	.074	-1.853	104.964
	3	47.998	15.334	.322	-9.782	105.777
	4	54.711(*)	13.483	.020	3.907	105.516
	5	37.317	13.147	.712	-12.221	86.855
	6	48.663	13.141	.061	-.852	98.178
	7	49.365	13.553	.073	-1.704	100.434
	8	59.009(*)	14.035	.013	6.125	111.893
	9	46.328	14.041	.199	-6.578	99.234
	10	19.608	16.971	1.000	-44.338	83.554
	11	16.955	14.810	1.000	-38.848	72.759
	12	-73.985	51.855	1.000	-269.374	121.404
	13	-55.695	29.995	1.000	-168.716	57.327
	14	-45.285	19.431	1.000	-118.502	27.931

Based on estimated marginal means
* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

H.5.2 Tukey HSD Post Hoc Pairwise Comparisons (Mean RT by Session)

Table H.5.2.1 Tukey HSD Post Hoc Pairwise Comparisons (Mean RT by Session)

Session	Mean RT	vs	Session	Mean RT	Difference	**p<.05**
1	252.94586778723700		2	247.01021808126700	5.93564970596984	ns
1	252.94586778723700		3	250.56782598080800	2.37804180642837	ns
1	252.94586778723700		4	243.85434855585500	9.09151923138171	ns
1	252.94586778723700		5	261.24869504182200	-8.30282725458551	ns
1	252.94586778723700		6	249.90304233716900	3.04282545006797	ns
1	252.94586778723700		7	249.20065141760800	3.74521636962896	ns
1	252.94586778723700		8	239.55652253524100	13.38934525199560	ns
1	252.94586778723700		9	252.23782514489200	0.70804264234462	ns
1	252.94586778723700		10	278.95782736073400	-26.01195957349680	ns
1	252.94586778723700		11	281.61021721881400	-28.66434943157690	ns
1	252.94586778723700		12	372.55065420399600	-119.60478641675900	ns
1	252.94586778723700		13	354.26021741784100	-101.31434963060500	ns
1	252.94586778723700		14	343.85108649212400	-90.90521870488700	ns
1	252.94586778723700		15	298.56565259850500	-45.61978481126860	ns
2	247.01021808126700		3	250.56782598080800	-3.55760789954147	ns
2	247.01021808126700		4	243.85434855585500	3.15586952541187	ns
2	247.01021808126700		5	261.24869504182200	-14.23847696055530	ns
2	247.01021808126700		6	249.90304233716900	-2.89282425590187	ns
2	247.01021808126700		7	249.20065141760800	-2.19043333634087	ns
2	247.01021808126700		8	239.55652253524100	7.45369554602576	ns

2	247.01021808126700	9	252.23782514489200	-5.22760706362521	ns
2	247.01021808126700	10	278.95782736073400	-31.94760927946670	ns
2	247.01021808126700	11	281.61021721881400	-34.59999913754670	ns
2	247.01021808126700	12	372.55065420399600	-125.54043612272900	ns
2	247.01021808126700	13	354.26021741784100	-107.24999933657400	ns
2	247.01021808126700	14	343.85108649212400	-96.84086841085690	ns
2	247.01021808126700	15	298.56565259850500	-51.55543451723840	ns
3	250.56782598080800	4	243.85434855585500	6.71347742495334	ns
3	250.56782598080800	5	261.24869504182200	-10.68086906101390	ns
3	250.56782598080800	6	249.90304233716900	0.66478364363960	ns
3	250.56782598080800	7	249.20065141760800	1.36717456320059	ns
3	250.56782598080800	8	239.55652253524100	11.01130344556720	ns
3	250.56782598080800	9	252.23782514489200	-1.66999916408375	ns
3	250.56782598080800	10	278.95782736073400	-28.39000137992520	ns
3	250.56782598080800	11	281.61021721881400	-31.04239123800530	ns
3	250.56782598080800	12	372.55065420399600	-121.98282822318700	ns
3	250.56782598080800	13	354.26021741784100	-103.69239143703300	ns
3	250.56782598080800	14	343.85108649212400	-93.28326051131540	ns
3	250.56782598080800	15	298.56565259850500	-47.99782661769690	ns
4	243.85434855585500	5	261.24869504182200	-17.39434648596720	ns
4	243.85434855585500	6	249.90304233716900	-6.04869378131374	ns
4	243.85434855585500	7	249.20065141760800	-5.34630286175275	ns
4	243.85434855585500	8	239.55652253524100	4.29782602061388	ns
4	243.85434855585500	9	252.23782514489200	-8.38347658903709	ns
4	243.85434855585500	10	278.95782736073400	-35.10347880487850	ns
4	243.85434855585500	11	281.61021721881400	-37.75586866295860	ns
4	243.85434855585500	12	372.55065420399600	-128.69630564814000	ns
4	243.85434855585500	13	354.26021741784100	-110.40586886198600	ns
4	243.85434855585500	14	343.85108649212400	-99.99673793626870	ns
4	243.85434855585500	15	298.56565259850500	-54.71130404265030	ns
5	261.24869504182200	6	249.90304233716900	11.34565270465350	ns
5	261.24869504182200	7	249.20065141760800	12.04804362421450	ns
5	261.24869504182200	8	239.55652253524100	21.69217250658110	ns
5	261.24869504182200	9	252.23782514489200	9.01086989693013	ns
5	261.24869504182200	10	278.95782736073400	-17.70913231891130	ns
5	261.24869504182200	11	281.61021721881400	-20.36152217699140	ns
5	261.24869504182200	12	372.55065420399600	-111.30195916217300	ns
5	261.24869504182200	13	354.26021741784100	-93.01152237601910	ns
5	261.24869504182200	14	343.85108649212400	-82.60239145030150	ns
5	261.24869504182200	15	298.56565259850500	-37.31695755668310	ns
6	249.90304233716900	7	249.20065141760800	0.70239091956100	ns
6	249.90304233716900	8	239.55652253524100	10.34651980192760	ns
6	249.90304233716900	9	252.23782514489200	-2.33478280772334	ns
6	249.90304233716900	10	278.95782736073400	-29.05478502356480	ns

6	249.90304233716900	11	281.61021721881400	-31.70717488164490	ns
6	249.90304233716900	12	372.55065420399600	-122.64761186682700	ns
6	249.90304233716900	13	354.26021741784100	-104.35717508067300	ns
6	249.90304233716900	14	343.85108649212400	-93.94804415495500	ns
6	249.90304233716900	15	298.56565259850500	-48.66261026133650	ns
7	249.20065141760800	8	239.55652253524100	9.64412888236663	ns
7	249.20065141760800	9	252.23782514489200	-3.03717372728434	ns
7	249.20065141760800	10	278.95782736073400	-29.75717594312580	ns
7	249.20065141760800	11	281.61021721881400	-32.40956580120590	ns
7	249.20065141760800	12	372.55065420399600	-123.35000278638800	ns
7	249.20065141760800	13	354.26021741784100	-105.05956600023400	ns
7	249.20065141760800	14	343.85108649212400	-94.65043507451600	ns
7	249.20065141760800	15	298.56565259850500	-49.36500118089750	ns
8	239.55652253524100	9	252.23782514489200	-12.68130260965100	ns
8	239.55652253524100	10	278.95782736073400	-39.40130482549240	ns
8	239.55652253524100	11	281.61021721881400	-42.05369468357250	ns
8	239.55652253524100	12	372.55065420399600	-132.99413166875400	ns
8	239.55652253524100	13	354.26021741784100	-114.70369488260000	ns
8	239.55652253524100	14	343.85108649212400	-104.29456395688300	ns
8	239.55652253524100	15	298.56565259850500	-59.00913006326420	ns
9	252.23782514489200	10	278.95782736073400	-26.72000221584140	ns
9	252.23782514489200	11	281.61021721881400	-29.37239207392150	ns
9	252.23782514489200	12	372.55065420399600	-120.31282905910300	ns
9	252.23782514489200	13	354.26021741784100	-102.02239227294900	ns
9	252.23782514489200	14	343.85108649212400	-91.61326134723170	ns
9	252.23782514489200	15	298.56565259850500	-46.32782745361320	ns
10	278.95782736073400	11	281.61021721881400	-2.65238985808008	ns
10	278.95782736073400	12	372.55065420399600	-93.59282684326190	ns
10	278.95782736073400	13	354.26021741784100	-75.30239005710770	ns
10	278.95782736073400	14	343.85108649212400	-64.89325913139020	ns
10	278.95782736073400	15	298.56565259850500	-19.60782523777170	ns
11	281.61021721881400	12	372.55065420399600	-90.94043698518190	ns
11	281.61021721881400	13	354.26021741784100	-72.65000019902770	ns
11	281.61021721881400	14	343.85108649212400	-62.24086927331010	ns
11	281.61021721881400	15	298.56565259850500	-16.95543537969170	ns
12	372.55065420399600	13	354.26021741784100	18.29043678615420	ns
12	372.55065420399600	14	343.85108649212400	28.69956771187170	ns
12	372.55065420399600	15	298.56565259850500	73.98500160549020	ns
13	354.26021741784100	14	343.85108649212400	10.40913092571750	ns
13	354.26021741784100	15	298.56565259850500	55.69456481933600	ns
14	343.85108649212400	15	298.56565259850500	45.28543389361850	ns

APPENDIX I. FACTOR ANALYSIS

I.1 Baseline Factor Analysis

Table I.1.1 Baseline MO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.587
Bartlett's Test of Sphericity	Approx. Chi-Square	89.236
	df	21
	Sig.	.000

Table I.1.2 Baseline Communalities(a)

	Initial
Baseline WOMBAT	.425
Baseline LNS	.199
Base LS_# crct	.290
Base NS_# crct	.845
Baseline MTT	.832
Base IGT good	.256
BAsePVT	.226

Extraction Method: Maximum Likelihood.
a One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Table I.1.3 Baseline Total Variance Explained

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.381	34.015	34.015	1.938	27.680	27.680
2	1.908	27.257	61.272	1.634	23.350	51.030
3	.922	13.175	74.448			
4	.705	10.075	84.522			
5	.579	8.272	92.795			
6	.420	6.001	98.795			
7	.084	1.205	100.000			

Extraction Method: Maximum Likelihood.

Factor Matrix(a)

a Attempted to extract 2 factors. More than 25 iterations required. (Convergence=.135). Extraction was terminated.

Table I.1.4 Baseline Rotated Factor Matrix(a)

	Factor	
	1	2
Baseline WOMBAT	.085	.905

Baseline LNS	-.021	.416
Base LS_# crct	-.116	.544
Base NS_# crct	.932	-.359
Baseline MTT	.862	-.275
Base IGT good	.370	.293
BAsePVT	-.411	-.236

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 3 iterations.

Table I.1.5 Baseline Factor Transformation Matrix

Factor	1	2
1	.932	-.363
2	.363	.932

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.

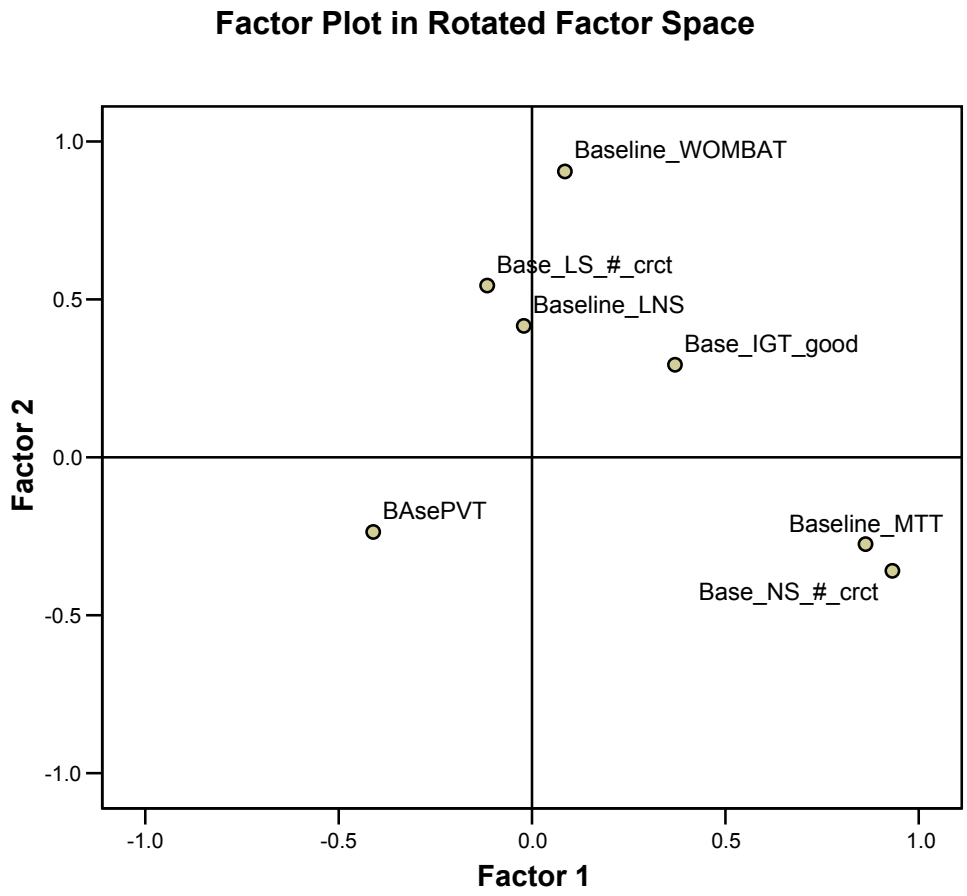


Figure I.1.1 Baseline Factor Plot Rotated in Factor Space

I.2 Sleep Deprived Factor Analysis

Table I.2.1 Sleep Deprived KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.649
Bartlett's Test of Sphericity	Approx. Chi-Square	44.276
	df	21
	Sig.	.002

Table I.2.2 Sleep Deprived Communalities(a)

	Initial
SD WOMBAT	.430
SD LNS	.237
SD LS_# crct	.332
SD NS_# crct	.238
SD MTT	.222
SD IGT good	.340
SDPVT	.172

Extraction Method: Maximum Likelihood.
a One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Table I.2.3 Sleep Deprived Total Variance Explained

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.411	34.438	34.438	1.755	25.066	25.066
2	1.403	20.049	54.487	1.115	15.932	40.997
3	.947	13.529	68.017			
4	.804	11.492	79.509			
5	.567	8.095	87.604			
6	.491	7.013	94.617			
7	.377	5.383	100.000			

Extraction Method: Maximum Likelihood.

Factor Matrix(a)

a 2 factors extracted. 20 iterations required.

Table I.2.4 Sleep Deprived Goodness-of-fit Test

Chi-Square	df	Sig.
5.006	8	.757

Table I.2.5 Sleep Deprived Rotated Factor Matrix(a)

	Factor	
	1	2
SD WOMBAT	.811	-.024
SD LNS	.437	.080
SD LS_# crct	.566	.019
SD NS_# crct	.248	.419
SD MTT	-.169	.892
SD IGT good	.627	.232
SDPVT	-.319	-.287

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 3 iterations.

Table I.2.6 Sleep Deprived Factor Transformation Matrix

Factor	1	2
1	-.248	.969
2	.969	.248

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.

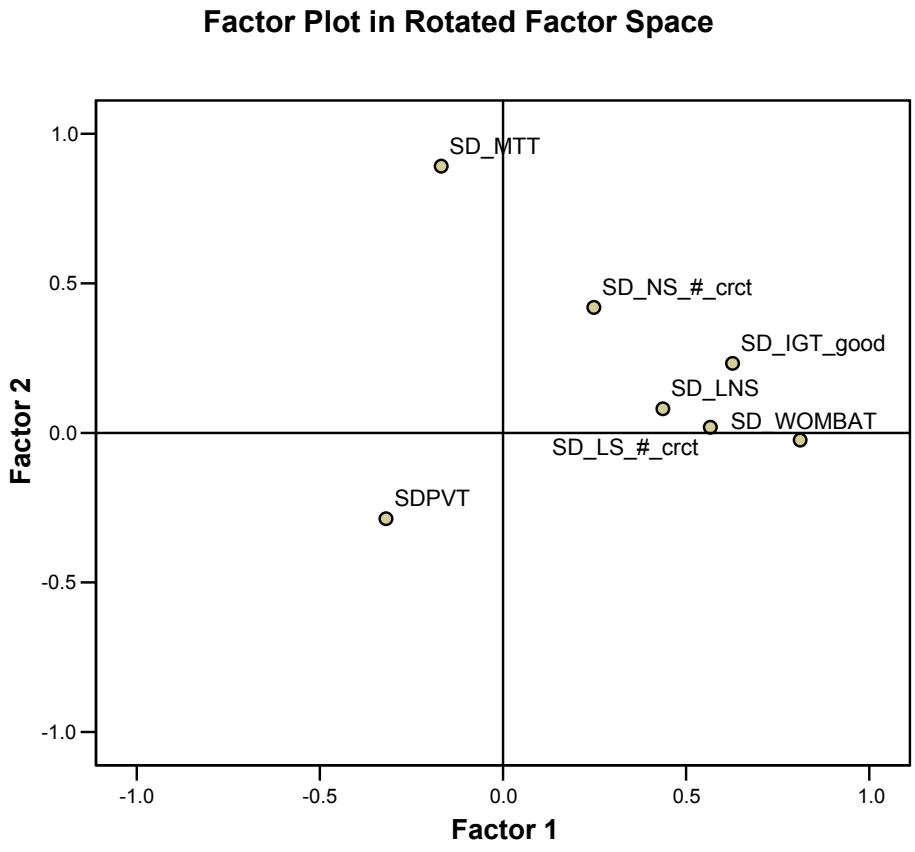


Figure I.2.1 Sleep Deprived Factor Plot Rotated in Factor Space

I.3 Difference Scores Factor Analysis

Table I.3.1 Difference KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.404
Bartlett's Test of Sphericity	Approx. Chi-Square	21.408
	df	21
	Sig.	.434

Table I.3.2 Difference Communalities(a)

	Initial
WOMBAT DIFF	.251
LNS Diff	.052
LS Diff	.100
NS Diff	.100
MTT Diff	.201
Good IGT Diff	.158
PVT diff	.092

Extraction Method: Maximum Likelihood.
a One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Table I.3.3 Difference Total Variance Explained

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.499	21.417	21.417	1.144	16.345	16.345
2	1.334	19.062	40.480	1.051	15.017	31.362
3	1.187	16.951	57.431	.578	8.250	39.612
4	.977	13.957	71.388			
5	.881	12.585	83.973			
6	.681	9.723	93.696			
7	.441	6.304	100.000			

Extraction Method: Maximum Likelihood.

Factor Matrix(a)

a 3 factors extracted. 5 iterations required.

Table I.3.4 Difference Goodness-of-fit Test

Chi-Square	df	Sig.
2.078	3	.556

Table I.3.5 Difference Rotated Factor Matrix(a)

	Factor		
	1	2	3
Good IGT Diff	.992	-.098	.067
NS Diff	.283	.109	-.140
WOMBAT DIFF	.245	.893	.375
MTT Diff	.056	-.466	.159
LS Diff	.045	.049	.448
PVT diff	.116	-.044	-.334
LNS Diff	.028	.101	-.275

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 4 iterations.

Table I.3.6 Difference Factor Transformation Matrix

Factor	1	2	3
1	.805	.518	.288
2	.585	-.775	-.241
3	.098	.362	-.927

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.

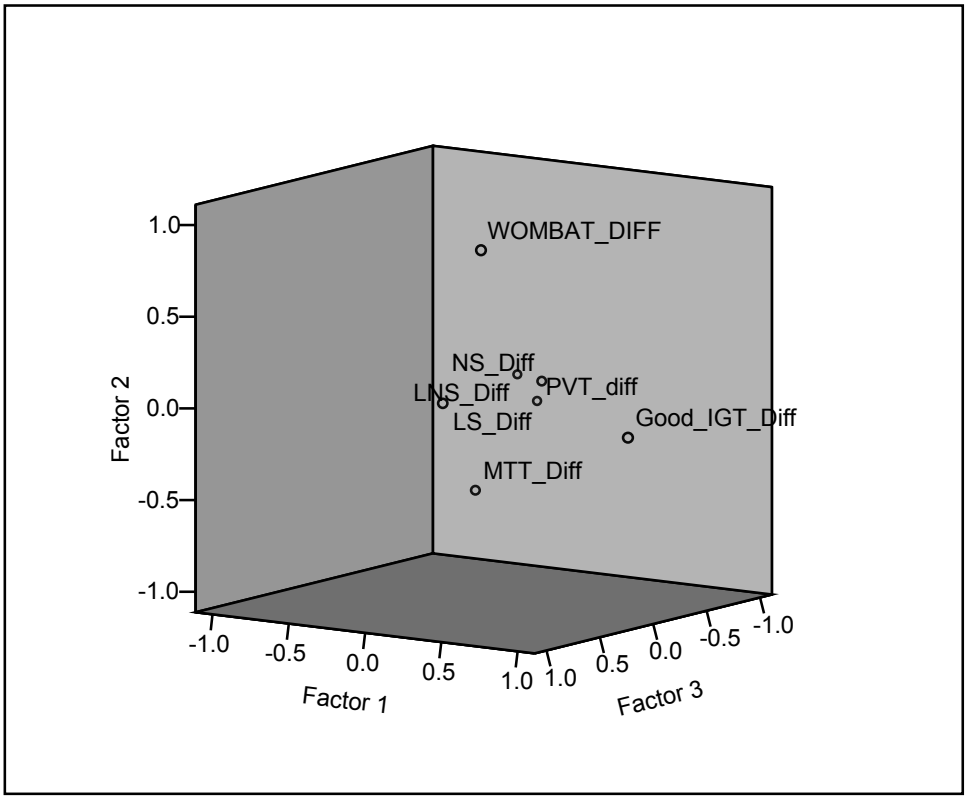


Figure I.3.1 Difference Factor Plot Rotated in Factor Space

VITA

Nancy L. Grugle

Nancy Grugle received a B.S. degree in Industrial and Manufacturing Systems Engineering from Ohio University in 1999. She went on to Virginia Polytechnic Institute and State University where she received an M.S. in Industrial and Systems Engineering and worked as an industrial engineer for the Army Research Laboratory during the summer of 2000. After graduation, Nancy worked as a Human Factors Engineer for Lockheed Martin Systems Integration in Owego, NY. In 2002, she returned to Virginia Polytechnic Institute and State University to pursue her Ph.D. in Industrial and Systems Engineering. Nancy conducted her dissertation research with Walter Reed Army Institute of Research on the effects of sleep deprivation on executive function and situation awareness. She will graduate in May, 2005 and begin work as an assistant professor of Industrial Engineering at Cleveland State University in July, 2005.